Development and Validation of a Regional Shelf Model for Maritime Canada based on the NEMO-OPA Circulation Model

D. Brickman and A. Drozdowski

Ocean and Ecosystem Sciences Division
Maritimes Region
Fisheries and Oceans Canada

Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, Nova Scotia
Canada B2Y 4A2

2012

Canadian Technical Report of Hydrography and Ocean Sciences 278

Fisheries and Oceans Canada | Pêches et Océans Canada

Canada

Canadian Technical Report of Hydrography and Ocean Sciences

Technical reports contain scientific and technical information of a type that represents a contribution to existing knowledge but which is not normally found in the primary literature. The subject matter is generally related to programs and interests of the Oceans and Science sectors of Fisheries and Oceans Canada.

Technical reports may be cited as full publications. The correct citation appears above the abstract of each report. Each report is abstracted in the data base *Aquatic Sciences and Fisheries Abstracts.*

Technical reports are produced regionally but are numbered nationally. Requests for individual reports will be filled by the issuing establishment listed on the front cover and title page.

Regional and headquarters establishments of Ocean Science and Surveys ceased publication of their various report series as of December 1981. A complete listing of these publications and the last number issued under each title are published in the *Canadian Journal of Fisheries and Aquatic Sciences*, Volume 38: Index to Publications 1981. The current series began with Report Number 1 in January 1982.

Rapport technique canadien sur l'hydrographie et les sciences océaniques

Les rapports techniques contiennent des renseignements scientifiques et techniques qui constituent une contribution aux connaissances actuelles mais que l'on ne trouve pas normalement dans les revues scientifiques. Le sujet est généralement rattaché aux programmes et intérêts des secteurs des Océans et des Sciences de Pêches et Océans Canada.

Les rapports techniques peuvent être cités comme des publications à part entière. Le titre exact figure au-dessus du résumé de chaque rapport. Les rapports techniques sont résumés dans la base de données *Résumés des sciences aquatiques et halieutiques*.

Les rapports techniques sont produits à l'échelon régional, mais numérotés à l'échelon national. Les demandes de rapports seront satisfaites par l'établissement auteur dont le nom figure sur la couverture et la page de titre.

Les établissements de l'ancien secteur des Sciences et Levés océaniques dans les régions et à l'administration centrale ont cessé de publier leurs diverses séries de rapports en décembre 1981. Vous trouverez dans l'index des publications du volume 38 du *Journal canadien des sciences halieutiques et aquatiques*, la liste de ces publications ainsi que le dernier numéro paru dans chaque catégorie. La nouvelle série a commencé avec la publication du rapport numéro 1 en janvier 1982.

Canadian Technical Report of
Hydrography and Ocean Sciences
2012

Development and Validation of a Regional Shelf Model for Maritime Canada based on the NEMO-OPA Circulation Model

by

David Brickman
and Adam Drozdowski

david.brickman@dfo-mpo.gc.ca
adam.drozdowski@dfo-mpo.gc.ca

Ocean and Ecosystem Sciences Division
Maritimes Region
Fisheries and Oceans Canada

Bedford Institute of Oceanography
P.O. Box 1006
Dartmouth, N.S.
Canada B2Y 4A2

©Her Majesty the Queen in Right of Canada 2008
Cat. No. Fs 97-18/240E ISSN 0711-6764

Correct citation for this publication:

Brickman, D., and A. Drozdowski 2012. Development and Validation of a Regional Shelf Model for Maritime Canada based on the NEMO-OPA Circulation Model.
Can. Tech. Rep. Hydrogr. Ocean Sci. 278: vii + 57 pp.

Table of Contents

List of Figures vi

1 Introduction 1

2 Model Details 3
2.1 Code changes . . . 5
2.1.1 I/O Routines . . . 6
2.1.2 Smagorinsky scheme . . . 7
2.1.3 Open Boundary Conditions (OBCs) . . . 7
2.1.4 Tides and related issues . . . 7
2.1.5 River input . . . 9
2.1.6 Tracer modules . . . 10
2.1.7 Particle tracking . . . 11
2.1.8 Other minor changes of note . . . 11
2.1.9 Closing remarks . . . 12

3 Model Setup 12
3.1 TS fields . . . 13
3.2 Open Boundary Fields . . . 13
3.3 Surface forcing . . . 14

4 Model Validation I – General 14
4.1 Validation Results . . . 17

5 Model Validation - II: Quantitative comparison to current meter and ice data 20
5.1 Current meter data distribution . . . 21
5.2 Velocity Comparison: Global Analysis . . . 25
5.3 Velocity Comparison: Seasonal Climatology . . . 33
5.4 Model Validation: Ice Model . . . 40

6 Summary 42

7 Acknowledgements 43

8 References 43

List of Figures

1 Schematic of the circulation in Atlantic Canada. The dashed black line is the model domain. For place names see figure 3. 2

2 MC domain with schematic of lateral boundary forcing. OBC forcing includes velocity, computed from the thermal wind relation (Vth), plus tidal forcing (SSH and transport). Analytic barotropic transports are added at 2 locations on the eastern boundary, balanced by an equivalent outflow at the western boundary shelf break. See text for more details. 8

3 MC domain including streamlines illustrating the interior flow and its connection to the valve forcing. Although based on actual particle tracks, the streamlines should be taken as schematic. The 3 interior sections indicated are, from W-E: Cape Sable Island (CSI), Halifax (HFX), and Cabot Strait (CS). The Halifax section contains 3 subsections: coast-240m isobath (blue), coast-340m (blue + green), and coast-1000m (blue+green+black). Abbreviations are: GSL=Gulf of St. Lawrence, SS=Scotian Shelf, GoM=Gulf of Maine, GB=Georges Bank, NEC=Northeast Channel, EB=Emerald Bank, WB=Western Bank, SIB=Sable Island Bank, BQ=Banquereau Bank, LC=Laurentian Channel, AC=Anticosti, SLE=St. Lawrence Estuary, SBI=Strait of Belle Isle, SB=shelf break. 15

4 Model transports versus data. (a) Model's net transport through Cabot Strait (shown for illustrative purposes). (b) Halifax section transport – coast to 240m isobath. Black line is Loder et al.(2003) estimate. Cyan boxes bound the estimates from Anderson and Smith (1989). (c) CSI transport. Black line is data from Smith (1983). 18

5 Development of ice concentration field. Left column is model, right is data. Scales are equivalent. 19

6 Comparison of model ice volume versus data. Red lines are monthly data, mean ± 1 std. Black lines are model: thick = monthly mean; thin = daily values. 20

7 Colour coded symbol plot of the total number of months of data in the 0-250m depth bin. 22

8 Histogram of the number of months of data in the 0-250m depth bin. . . 23

9 Spatial symbol plot of the number of distinct years of data in the 0-250m depth bin. 24

10 Histogram of the number of distinct years of data in the 0-250m depth bin. (NB: dlon, dlat refer to half grid cell dimension.) 25

11 Top: Histogram of the fraction of locations that have N distinct months of data (for the 0-250m depth bin). Btm: Fraction of locations with $> N$ distinct months of data. (NB: dlon, dlat refer to half grid cell dimension.) 26

12 Illustration of vector differences. The data vector is in red. d1 and d2 are model–data difference vectors. The angles o_1 and o_2 are angle errors with sign defined by the right hand rule: data × model 27
13 Scatterplot of model speed versus observation speed (m/s). Note the log scale. The model equals data line is drawn through the data. 29
14 Frequency distribution of model–observation differences for monthly mean speeds. 30
15 Frequency distribution of the ratio of model to observed speed on a log scale. 30
16 Magnitude of model–observation vector differences (top) and error angle (bottom) 31
17 The distribution of speeds for both model and observation. 32
18 Horizontal distribution of data. Boxes define regions used for the analysis 36
19 Model Skill for various regions and layers – weak selection criteria. 37
20 Number of data associated with the skill table. 38
21 Comparison of model skill for weak versus strong criteria for the Southern GSL. CS and Hfx boxes. including the data counts. 39
22 Definition of Gilbert Boxes (reproduced from Galbraith et al. 2011) . . . 40
23 Time series of 49 years of ice conditions in Gilbert Box 1. Horizontal line denotes the CANOPA normal year prediction. 46
24 Time series of 49 years of ice conditions in Gilbert Box 2. Horizontal line denotes the CANOPA normal year prediction. 46
25 Time series of 49 years of ice conditions in Gilbert Box 3. Horizontal line denotes the CANOPA normal year prediction. 47
26 Time series of 49 years of ice conditions in Gilbert Box 4. Horizontal line denotes the CANOPA normal year prediction. 47
27 Time series of 49 years of ice conditions in Gilbert Box 5. Horizontal line denotes the CANOPA normal year prediction. 48
28 Time series of 49 years of ice conditions in Gilbert Box 6. Horizontal line denotes the CANOPA normal year prediction. 48
29 Time series of 49 years of ice conditions in Gilbert Box 7. Horizontal line denotes the CANOPA normal year prediction. 49
30 Time series of 49 years of ice conditions in Gilbert Box 8. Horizontal line denotes the CANOPA normal year prediction. 49
31 Time series of 49 years of ice conditions in Gilbert Box 10. Horizontal line denotes the CANOPA normal year prediction. 50
32 Time series of 49 years of ice conditions in the entire Gulf of Saint Lawrence. Horizontal line denotes the CANOPA normal year prediction. 50
33 Time series of 49 years of regionally averaged (weighted by area) ice conditions in the Gulf of Saint Lawrence. 51

34 Regional 49 year mean of first appearance and CANOPA normal year value. Included are lines showing standard deviation departure from the mean of the data. 51
35 Regional 49 year mean of last appearance and CANOPA normal year value. Included are lines showing standard deviation departure from the mean of the data. 52
36 Regional 49 year mean of peak volume and CANOPA normal year value. Included are lines showing standard deviation departure from the mean of the data. 52
37 Regional 49 year mean of normalized peak volume and CANOPA normal year value. Included are lines showing standard deviation departure from the mean of the data. All values normalized by peak volume of data. . . 53
38 Monthly mean peak ice volume based on 49 years of data and CANOPA normal year. Included are lines showing standard deviation departure from the mean of the data. Part 1: Sum of all regions and western/north-western regions. 54
39 Monthly mean peak ice volume based on 49 years of data and CANOPA normal year. Included are lines showing standard deviation departure from the mean of the data. Part 2 Eastern and Central Regions 55
40 Monthly mean ice area based on 49 years of data and CANOPA normal year. Included are lines showing standard deviation departure from the mean of the data. Part 1: Sum of all regions and western/north-western regions. 56
41 Monthly mean ice area based on 49 years of data and CANOPA normal year. Included are lines showing standard deviation departure from the mean of the data. Part 2 Eastern and Central Regions 57

Abstract

Brickman, D., and A. Drozdowski 2012. Development and Validation of a Regional Shelf Model for Maritime Canada based on the NEMO-OPA Circulation Model.
Can. Tech. Rep. Hydrogr. Ocean Sci. 278: vii + 57 pp.

This report describes the development and validation of a NEMO-OPA numerical circulation model ("CANOPA" model) specifically configured on a geographic domain that is relevant to ecosystem applications of interest to DFO researchers in the Quebec, Gulf and Maritime regions of Eastern Canada. The NEMO-OPA ocean model was originally developed for global ocean application, and part of the motivation for this report is the description of the modifications enacted to adapt the model for shallow water purposes. The model code, changes made by the authors related to its use as a shelf model, and model set up are described. The philosophy behind model validation and model validation results are presented, including quantitative comparisons between model currents and current meter data, and the model's ice simulation versus 49 years of ice data.

Résumé

Brickman, D., et A. Drozdowski, 2012. Création et validation d'un modèle de plateau régional pour le secteur des Maritimes, au Canada fondé sur le modèle de circulation NEMO-OPA. Rapp. tech. can. hydrogr. sci. océan. 278 : vii + 57 p.

Le présent rapport décrit la création et la validation d'un modèle de circulation NEMO-OPA (modèle ≪CANOPA≫), conçu spécialement en fonction d'un secteur géographique pertinent aux applications liées aux écosystmes d'intérêt aux yeux des chercheurs du MPO dans les régions du Québec, du Golfe et des Maritimes de l'Est du Canada. À l'origine, le modèle océanographique NEMO-OPA a été conçu pour l'application océanographique globale et une partie de la motivation stimulant ce rapport correspond à la description des modifications apportées pour adapter le modèle aux eaux peu profondes. Le code du modèle et les changements apportés par les auteurs quant à son utilisation à titre de modèle de plateau, en plus de la configuration du modèle, y sont décrits. La philosophie appuyant la validation du modèle et les résultats de validation du modèle sont présentés, notamment les comparaisons quantitatives entre les courants du modèle et les données de mesure du courant et les comparaisons entre la simulation du mouvement des glaces du modèle et les données sur les glaces enregistrées sur 49 ans.

1 Introduction

This report describes the adaptation of a NEMO-OPA numerical ocean circulation model for application(s) on the shelf seas of Maritime Canada. The Nucleus for European Modelling of the Ocean – Ocean Parallisé (NEMO-OPA) ocean model (see www.nemo-ocean.eu) was adopted ca. 2006 as the one of the main circulation models at DFO, and its use is promoted by the DFO Centre for Ocean Model Development for Application (COMDA) Centre of Excellence (COE). The model was originally developed for deep (e.g. global) ocean application, and part of the motivation for this report is the description of the modifications enacted to adapt the model for shallow water purposes. The main motivation for this report is to provide details and output of a model specifically configured on a geographic domain that is relevant to ecosystem applications of interest to DFO researchers in the Quebec, Gulf and Maritime regions of Eastern Canada.

Circulation background:

The circulation along the eastern seaboard of Canada is characterized by a general southward flow of subpolar (cold, fresh) water from the Labrador Sea, and a northward flow of subtropical (warm, salty) water from the Gulf Stream (figure 1). The colder, fresher water from the Labrador Sea follows two pathways into the Maritime Canada region. Inshore Labrador shelf water enters the Gulf of St. Lawrence (GSL) through the Strait of Belle Isle (SBI). Labrador slope water flows along the shelfbreak toward the tail of the Grand Banks at which point it interacts with the northeastward flowing Gulf Stream.

Part of the flow along the Newfoundland shelf flows northward through Cabot Strait and does a loop through the GSL before exiting through Cabot Strait. The majority of the shelfbreak flow follows bathymetric contours, but exhibits incursions into the Laurentian Channel, the central Scotian Shelf (SS), and the Northeast channel. The area south of the SS/GoM (GoM = Gulf of Maine) is a region of high variability shared by a cold water recirculation gyre and the Gulf Stream. Meanders and warm-core rings shed by the Gulf Stream affect water properties in the outer SS/GoM shelf area, with an influence that decreases from west to east.

The seas of Maritime Canada are influenced by numerous river inputs, the most dominant being the outflow of the St. Lawrence Estuary which provides an annual average of about 15mSv of freshwater (>10 times the next highest inflow rate). This pulse of low salinity water peaks in April in the northern GSL, and can be traced as it flows southward through Cabot Strait and onto the Scotian Shelf where it arrives at Halifax in late summer. The volume flux through the Strait of Belle Isle is more than 10 times greater than that from the Estuary, and wends its way around the GSL before also passing through Cabot Strait. Thus to first order, the net (southward) transport through Cabot Strait is equal to the inflow at the Strait of Belle Isle. Part of the outflow through CS follows the coastline as

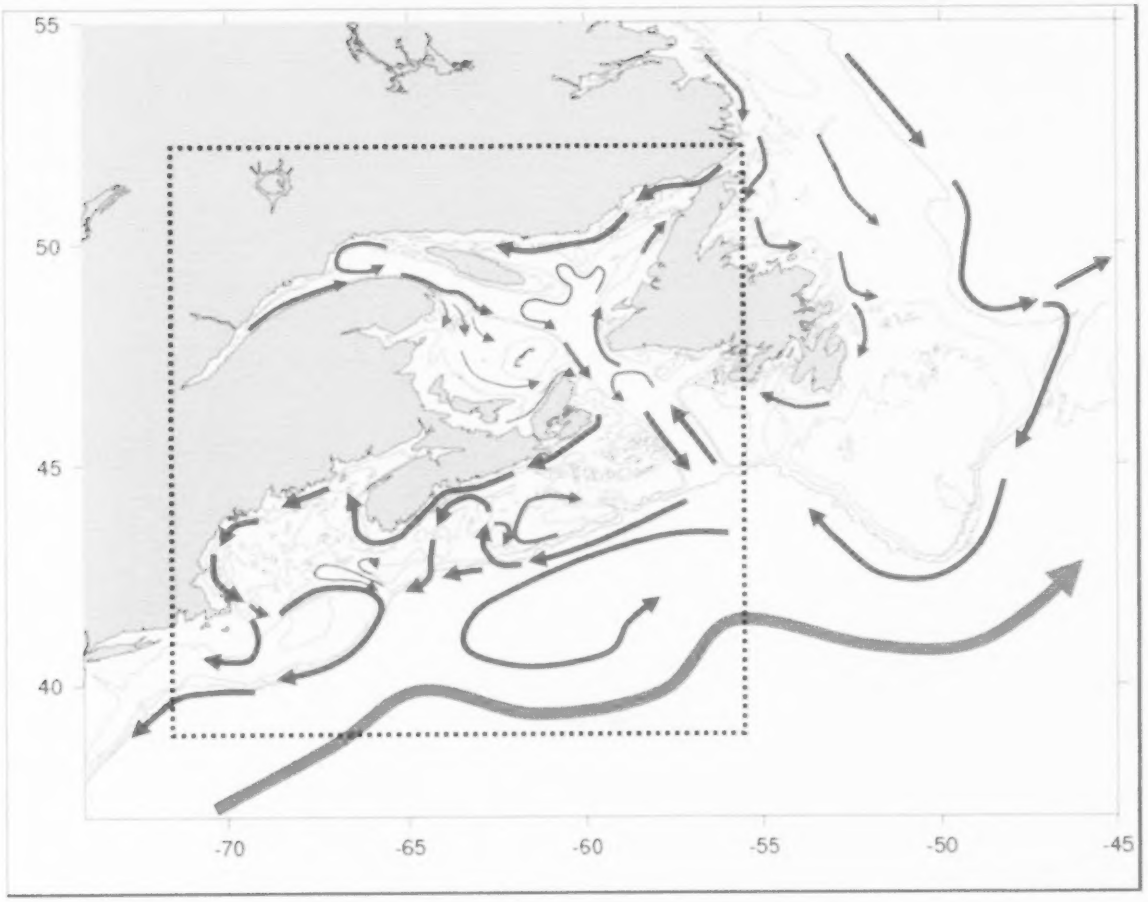


Figure 1: Schematic of the circulation in Atlantic Canada. The dashed black line is the model domain. For place names see figure 3.

the Nova Scotia current which eventually feeds the GoM coastal current. Another part travels southward along the western slope of the Laurentian Channel and becomes part of the SS shelfbreak current (see also figure 3).

From the above, one can see that the Gulf of St. Lawrence, Scotian Shelf, and Gulf of Maine form an interconnected set of shelf seas in Maritime Canada, and it is logical to model them using a single domain. As well, external (deep ocean) influences are much greater for the SS/GoM area than the GSL (whose primary input is through the SBI). This supports the view that the GSL acts more like a semi-enclosed sea (i.e. more influenced by surface fluxes) than the SS/GoM whose properties depend more on the interaction between the southward flow of subpolar water around the tail of the Banks and the northeastward flow of the Gulf Stream. From a modelling perspective this means that knowledge of the open boundary conditions (transport and TS properties)

are more important for the (boundary) area that affects the SS/GoM than the GSL. The choice of model domain is also motivated by the need to provide modelling support for various regional DFO research initiatives such as AZMP, the Aquatic Climate Change Adaptation Services Program, the Climate Change Science Initiative and the Maritime and Gulf region Ecosystem Research Initiatives.

The response of an ocean model is largely determined by its surface forcing. With the creation of various atmospheric reanalysis programs, principally NCEP (National Center for Environmental Prediction) and ECMWF (European Centre for Medium-range Weather Forecasts), forcing data for extended periods of time have now become available to drive ocean models. The NCEP reanalysis period 1958-2000 has been analyzed to produce the Common Ocean-ice Reference Experiments (CORE) normal year (NY) of forcing (Large and Yeager, 2004). This 365 day "cyclical year" dataset, which includes moving weather systems, is designed to represent the climatological atmospheric forcing with its climatological variability. The results reported in this document are based on the ocean model's response to CORE-NY forcing field variables.

The outline of this report is the following: Section 2 provides an overview of the NEMO-OPA model and details of the code changes incorporated by the authors. Section 3 provides more specific details of the model set up and creation of model input files. Section 4 describes details of the model validation exercise and some general validation results. Section 5 contains quantitative comparisons between model currents and current meter data, and the model ice simulation and ice data. The last section is a summary.

2 Model Details

The OPA model is a 3D, non-linear, hydrostatic, C-grid, primitive equation numerical ocean model written in parallelized fortran 90. It is the main model used by the NEMO group, and is currently referred to as NEMO-OPA. The majority of uses of the model are for open, or deep, ocean applications, with limited reported use as a regional, or shelf model. Indeed, at the time of writing the authors were unable to find any peer-reviewed published manuscripts reporting the use of NEMO-OPA as a shelf model. The model reported herein is based on OPA 9.0, with the incorporation of various updated routines since that release date (ca. 2005), as well as routines written by NEMO modellers at BIO (detailed below). For this reason, and in the spirit of the CANDIE Model (Sheng et al., 1998), we call our model the CANadian OPA or "CANOPA" model.

The governing equations of the model are standard so are not included here. (For more details see the NEMO manual (Madec 2008), hereafter refered to as "NEMO manual"). However, notable is the use of the "vector invariant" form of the primitive equations in which the momentum equation is written as

$$\frac{\partial V_h}{\partial t} = \{(\nabla \times V) \times V + \frac{1}{2}\nabla(V^2)\}_h - f\hat{k} \times V_h - \frac{1}{\rho_0}\nabla_h P + D + F \qquad (1)$$

where V is the velocity vector, the subscript h denotes the horizontal component, D is the diffusion operator, F is surface forcing, and the other symbols have their usual meaning. Numerous options are available for the free surface evolution equation (i.e. barotropic mode) of which we use the time-splitting version as we are interested in accurately modelling tidal motions.

The model is coded using generalized orthogonal curvilinear coordinates in the horizontal, although in practice the shape factors tend to be based on the usual spherical polar coordinate formulation. No GUI-based grid generating package exists for the NEMO-OPA model, as exist for the ROMS and POM models (e.g. Seagrid, see http://woodshole.er.usgs.gov/operations/modeling/seagrid/seagrid.html) and the Quoddy finite element model (Chaffey and Greenberg, 2003). This is considered a weakness for the NEMO modelling system, especially as more shelf-based applications are pursued. Author Drozdowski has adapted a version of the MATLAB Seagrid routine for use in the CANOPA system, but this effort is ongoing at the time of writing.

The model is setup on a grid that includes the GSL, SS, and GoM – the "Maritime Canada" (MC) domain (figure 1). The grid was derived as a subset of a NEMO tri-polar global spherical coordinate domain, so the grid lines run almost along latitude/longitude circles. The resolution is nominally 1/12-deg, which translates into grid cell dimensions ranging from $\sim$ 7-5km south-to-north. The MC domain has 234 cells in the south-north ("j") direction and 197 cells in the west-east ("i") direction, and spans a longitude/latitude range of roughly 72-55W to 38-52N.

Bathymetry for the model was derived from a combination of sources, predominately Atlantic Geoscience Centre (AGC) data (available at BIO), etopo2 (ETOPO2v2), and data provided by Dr. J. Chasse (DFO Gulf Region). These datasets were interpolated onto a regular $1/30 \times 1/30$ degree grid (spanning 77-38W to 35-57N) using an optimal interpolation routine. The model bathymetry at the grid centres (T-points) were interpolated from this regular 1/30 degree grid. Smoothing of the model bathymetry (in ij-space) was done using a simple boxcar smoother, in order to eliminate rough, "noisy", areas and produce a bathymetry more commensurate with the nominal 6km×6km grid cell size.

NEMO-OPA has multiple options for the vertical coordinates including sigma, z-level with or without partial cells, and a bottom boundary layer routine. The original CANOPA code was ported from a standard deep ocean application that used an equation for the z-levels (see NEMO manual):

$$z(k) = h_{sur} + h_0 * k - h_1 * \log(\cosh((k - h_{th})/h_{cr}))$$

where k is the vertical level and the $h_{...}$ are factors computed from input parameters. The above defines a vertical coordinate with level thickness that increases with k. For compatibility reasons CANOPA has kept the original 46 z-levels it inherited from the deep ocean application. This results in w-levels at $\sim$ (0, 6, 13, 20, 28, 37, 47, 58, 71, 85, 103, 123, 147, 175, 209, 248, 295, 350, ...) meters. Although suitable for most model

applications to date, these levels would be considered a bit coarse for a shelf model, in particular in the top 100m. A change in the vertical coordinate levels is a high priority for future model updates. However, note that this is potentially non-trivial as the model domain includes the upper Bay of Fundy in which tidal amplitudes can be O(5m) so that the thickness of the top layer may ultimately be dictated by this consideration.

The temperature (T) and salinity (S) input fields required by the model were created from data available from the DFO hydrographic climate database. All available coincident TS data for a region larger than the model domain were extracted from the database and then an optimal interpolation technique used to create monthly TS fields on a regular 0.25×0.25 degree grid, with vertical levels at (0,10,...,50,75,...,200,250,...,600,700,...,1200,1400, ...,2000,2500,...,5000) metres. These monthly fields were then interpolated onto the CANOPA MC grid to produce 12 files used as initial value and interior restoration fields for the model (see later).

The code contains numerous choices for both horizontal and vertical advection and diffusion schemes, selected using a combination of preprocessor *#Define* keys and variables input at runtime. In practice only the total variation diminishing (TVD) 3D advection scheme for T&S was found to work consistently (routine *traadv_tvd.F90*). For momentum, the combined energy/enstrophy conserving scheme was used to compute the vorticity term (see (1)). Horizontal diffusion, for tracer and momentum, was accomplished using a Laplacian scheme with coefficients computed using a Smagorinsky routine (see below). An implicit vertical mixing scheme (tracer and momentum) was selected with coefficients computed using the TKE routine. The latter is a version of a Mellor-Yamata level 1.5 scheme (Mellor and Yamata, 1982), and was chosen as it integrated considerably faster than the NEMO KPP scheme (Large et al., 1994). Vertical convection was accomplished using an enhanced vertical diffusion scheme with coefficient set to $5m^2/s$ when the water column was unstable.

The timestep for the baroclinic mode was 480s, with the barotropic timestep set to 8s. The model uses a leap frog time-stepping scheme with the Asselin time filtering parameter ($aftp$) set to 0.1 in the (e.g.) equation $ub = un + atfp * (ub + ua - 2un)$ (b, n, a denote before, now, and after variables).

CANOPA uses the LIM2 ice model coupled to the ocean model every 5 ocean timesteps. Details of the ice model can be found in Madec et al. (1998).

2.1 Code changes

A number of changes have been made to the code locally, notably to the Input/Output routines, plus others related specifically to the use of the model for shelf applications. Examples of the latter are the open boundary and tidal routines, a Smagorinsky scheme for horizontal mixing (Smagorinsky, 1963), and the addition of a river input module to allow for internal fluxes of fresh water which are an important part of the circulation in

Maritime Canada waters. We discuss the modifications below. The CANOPA code was also compared to the Global Ocean-Atmosphere Prediction and Predictability (GOAPP) code (see http://www.goapp.ca/index.php) obtained from Dr. F. Dupont (DFO) ca. 2008. While small differences were found and incorporated, few substantial differences were found in key modules pertaining to model performance. Thus the CANOPA code can be considered as up-to-date as the GOAPP version.

2.1.1 I/O Routines

The NEMO-OPA code is mostly based on NetCDF input and output files (re NetCDF see http://www.unidata.ucar.edu/software/netcdf/), supplemented by a few (albeit important) text (namelist) files. In a multiprocessor environment processors can share input files (i.e. can read the same file) but writing to the same file by more than one processor can lead to unstable results (seemingly operating system dependent). One of the changes in the CANOPA code was remedying all instances where runtime information was not output solely by the "head node" (processor 1 or the *lwp* variable).

The NEMO-OPA code also writes one NetCDF file per processor, a method that complicates directory structure management (due to the excessive number of files produced), and requires any postprocessing of model output to re-assemble (merge) the local processor arrays into the global domain array(s). This results in added analysis time and/or disk space requirements (the latter due to the fact that merging the N-processor output files requires twice the disk space, at least until the original N files are deleted). The choice of one output file per core is simple and may have been due to file size limitations in early releases of the NetCDF library. A major modification to the CANOPA code was designed and implemented by Dalhousie University computer science co-op student Chris Nickerson (2007-2008) wherein he re-coded the model to produce one output file per set of variables, instead of the N files. At the risk of oversimplification, this was achieved using looped MPP *send/receive* commands that moved local processor data to the head node which was responsible for output. During the same time period, changes to the basic set of variables in the output files, and file names, were made by authors Brickman and Drozdowski in order to simplify their analysis routines. The main result is that all regularly used model output is contained in 2 files: an "icemod" file containing the ice model variables, and an "aveTSUV"file which contains the rest of the model variables. Note that model output is time-averaged over the period between outputs. Because of these latter changes away from the NEMO-OPA structure, we did not release the new output module to the main NEMO consortium. (Note that a copy of the Nickerson routines (*lib_ncdf.F90* module) – easily modifiable – can be obtained by contacting author Brickman.)

2.1.2 Smagorinsky scheme

The authors adopted the Smagorinsky scheme for horizontal mixing coefficients written by Dr. Z. Wang (BIO) and subsequently made available to the NEMO-OPA community. The Smagorinsky algorithm (Smagorinsky, 1963) computes space and time dependent mixing coefficients (Am, Ah) based on the shear/strain in the horizontal velocity:

$$Am \sim \alpha * \sqrt{(\partial u/\partial x)^2 + (\partial v/\partial y)^2 + 0.5 * (\partial v/\partial x + \partial u/\partial y)^2}$$

where α is a parameter chosen as 0.1, and the tracer diffusivity Ah is related to the momentum diffusivity Am by a Prandt number: $Pr = Ah/Am = 0.1$

2.1.3 Open Boundary Conditions (OBCs)

For a regional ocean model the forcing at the open boundaries can play a major role in the solution in the interior. This is particularly true for the MC domain as it has 2 major boundary inflow regions (SBI and the shelfbreak south of Newfoundland) and 1 key outflow region at the shelfbreak along the western boundary (the "valves", figure 2) . Principal open boundary variables are velocity (or transport), TS, and surface elevation. The original NEMO code had provisions for OBCs but the module required modification to suit the specific requirements of the MC domain. The latter included code to allow a barotropic transport (and TS perturbations) to be added to regions of the open boundaries (specifically at the valves); a routine to calculate the barotropic open boundary transport based on the vertical integral of the baroclinic velocities; and a routine to calculate the boundary surface elevation based on the surface velocity and geostrophy. These 3 routines guarantee consistency between the open boundary velocities and surface elevations. They are complemented by a routine that computes the net transport through the boundaries and adjusts the boundary velocities in order to conserve volume inside the domain. This is necessary because there is no guarantee that the net boundary transport is zero, and systematic imbalances can lead to appreciable changes to the mean interior surface elevation.

2.1.4 Tides and related issues

The model code was modified to run a full set of tidal constituents, input as open boundary elevations and transports. Tidal forcing data were derived from a run of a global tidal model (Lyard et al., 2006). For the shelf region of Maritime Canada the 5 major constituents (M2, N2, S2, K1, O1) are the main ones (see Dupont et al., 2002, table 13) so these are the only ones for which forcing data were extracted.

Tidal elevation and barotropic velocity enter the model as boundary values (via the open boundary code, and see figure 2), called by the surface pressure gradient solver. Note that

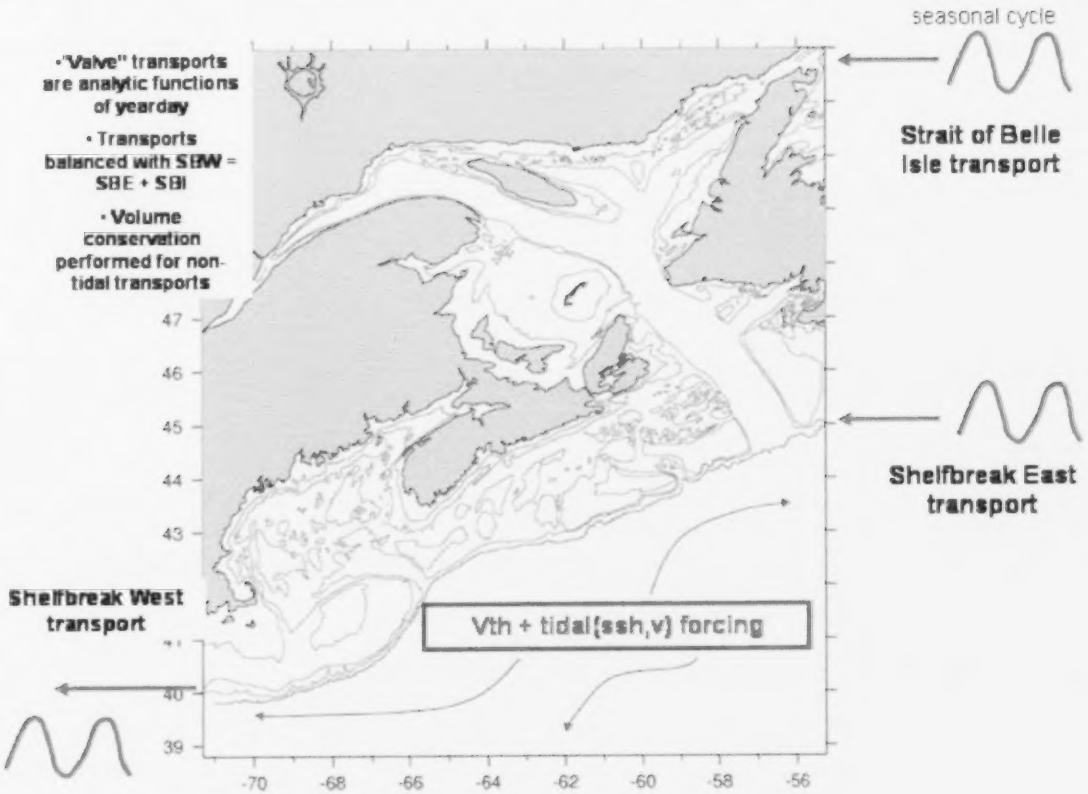


Figure 2: MC domain with schematic of lateral boundary forcing. OBC forcing includes velocity, computed from the thermal wind relation (Vth), plus tidal forcing (SSH and transport). Analytic barotropic transports are added at 2 locations on the eastern boundary, balanced by an equivalent outflow at the western boundary shelf break. See text for more details.

no body force was found necessary for tidal simulations. The Flather radiation condition (Flather, 1976) is used for the barotropic mode, which radiates (normal barotropic) velocity and sea surface height based on the difference between the applied boundary surface height values and values 1 cell adjacent to the boundary. Note that the volume conservation algorithm was **not** applied to the tidal component of open boundary transport as it was found that doing so eliminated most of the interior tidal energy. A result of the open boundary radiation condition is that the non-tidal transport, input at the open boundary (as OBCs, see later), may not be perfectly realized. This is because an applied transport can set up a sea level difference near the open boundary that results in an opposing flow due to the radiation algorithm. This opposing flow was found to be small, less than 10% of the applied transport. Note that this model response should not be unique to the CANOPA code but rather a feature of all limited area ocean models

that use a radiation boundary condition.

For baroclinic motions near the boundaries, the radiation and sponge layer algorithms were tested but ultimately turned off as they resulted in no obvious improvements to model performance. The open boundary algorithms for baroclinic motions will be re-investigated when migration to more up-to-date NEMO-OPA code is accomplished.

Running the model with tidal forcing complicates the determination of the time-mean flow quantities that we are typically interested in. This is due to the large tidal velocities in many regions of the domain. One solution is to output model variables at high frequency and then perform a tidal analysis to filter out the tidal components. This is impractical due to excessive disk space requirements. Recalling that the model is coded to produce time-averaged output, it is possible that the tidal signals will be averaged out if the averaging period is longer than the predominate tidal period(s). For example, since the 5 main tidal constituents have periods roughly 12 and 24 hours it would be expected that daily-averaged velocity fields would contain little tidal "biasing". Simulations using 5 tides and 1-5 day averaging periods showed that this tidal biasing effect can be as high as $\sim$ 5mm/s in absolute value. This can be large in areas where the mean is very small, but was found to be generally 1-5% in areas where there are noticeable currents. To avoid this problem, a routine that averages model velocity over the last M2 tidal period of each day was coded in order to provide accurate daily mean velocity fields. For the GoM, Bay of Fundy and SS, the M2 tidal energy is *at least* 7 times greater than the **sum** of the next 4 highest constituents, with this ratio falling to $\sim$ 2.4 for the GSL (From Dupont et al., 2002, table 13 the ratios are: GoM=12.9, BoF = 13.3, SS=7, GSL=2.4). Thus it is reasonable to run only the M2 tide in this region, especially when one is interested in mean velocity fields. While neither of these options is perfect, both are considered acceptable ways to run the model on the MC domain if the focus is on mean circulation properties. Most of the runs reported in this document used only the M2 tide, and the M2-tidal averaging routine.

2.1.5 River input

One of the main modifications in CANOPA is the incorporation of a river inflow module. The original NEMO-OPA code allows river input as an evaporation-minus-precipitation ($E - P$) input at the river mouth. While this may work well for some rivers, or in low resolution modelling, it was found that for the MC domain this technique did not work. River inflow is essentially an interior open boundary problem in which velocity (or transport), and T&S are applied at the boundary between land (dry cells) and ocean (wet cells), thus "opening" this boundary. A key difference between the above and the generic OBC code is that the latter involves only the edges of the domain (cells easily identified and owned by a single tile), while the former involves interior cells that may be shared by multiple tiles.

The MC domain has 78 rivers that flow into it. Of the roughly 25mSv annual average

inflow, about 15mSv comes from the St. Lawrence Estuary (3 rivers) with the St. John river (~1mSv) having the 4th largest contribution. Monthly timeseries of river transport from 1948 to the present were prepared by Dr. J. Chasse (DFO Gulf region, *in preparation*). Note that the data for flow past Quebec City (~12mSv) is available on the DFO climate database. The model inputs the data from a text file (Rivers.dat) that contains the grid (i, j) of the river mouth wet-cell, the cell-face through which the flux enters the grid cell, and the timeseries data.

The CANOPA rivers code is contained in the rivers module *rivers.F90*, activated by the key *key_RIVER_INPUT*. The process is initiated in *step.F90* by a call to *riv_init* on the first timestep. This routine reads Rivers.dat, and assigns the local processor logical array *do_river* based on whether the upstream river cell (i.e. the river's associated dry land cell) lies within the bounds of a given processor's tile. This guarantees no array out-of-bounds can occur, and importantly allows possible sharing of a given river between tiles, a feature that (after extensive debugging) was found to be crucial to the success of the river algorithm. Isolating a given river to a single processor did not work for all layout tilings tested. This was determined to be due to the fact that some intermediary computations near the halo-zones, in the no-sharing case, were not carried out for adjacent tiles, which can lead to problems when the local processor array halo-zone linking routine *lbc_lnk* is called by related subroutines (essentially the "correct" computation being overwritten by a random or zero value from the adjacent tile). The *riv_init* subroutine also initializes a *rivmask* array that sets upstream river dry cells to "wet". This array replaces the *tmask* variable in certain computations in the TVD advection scheme.

Every timestep *step* calls *riv_tra* and *riv_dyn*. The former routine sets the S at the upstream river cell to FW_SALT (a shared parameter usually equal 1) and the T at the upstream river cell to the value at the river mouth satisfying a zero gradient criterion. The latter routine retrieves the river transport and computes and sets the baroclinic velocity through the relevant cell-face. These 2 routines serve to set the correct values for the subsequent call to the advection scheme.

The surface pressure solver calls *dyn_ts* which sets the barotropic velocity across the river cell-face. This was found to be crucial in getting the freshwater transport to sensibly move "downstream" away from the river mouth as would be expected. Finally, routine *dyn_nxt*, which updates the after velocity using the leap-frog scheme, calls *riv_dyna* which sets the after velocity for the river inflows so that this information is correctly passed to the leap-frog scheme and subsequently available to the horizontal divergence algorithm. The correct horizontal divergence at the river mouth is essential as it allows the vertical velocity to be computed in an internally consistent way.

2.1.6 Tracer modules

The inherited NEMO-OPA model contained code for the evolution of passive and active tracers (or BioGeoChemical Models – BGCMs). Functionality of this code ranged from

workable for the passive tracer code, to uncompilable for various BGCMs. In general this code was poorly structured and lacked documentation.

Authors Brickman and Drozdowski modified the code to create 1 basic self-contained module for (each of the) passive and active tracers, structured around a controlling "*_model.F90" file and a "*_lib_model.F90" file that contains all the relevant subroutines for the controlling model. At the time of writing the CANOPA code contains a generic passive tracer module, and a simple "NPZ" BGCM module. The passive tracer module has been ported (by author Drozdowski) to the NEMO Arctic model developed and run at BIO (by Dr. Y Lu and others). The BGCM module has been modified (by the authors and Dr. D. Lavoie, IML) to run the IML "LifeMaker" model (LeFouest et al. 2005) on the MC domain, although this is still in the development stage.

The above modules can also run in offline mode, whereby the velocity and T&S fields are input from a previous run and the tracer module is stepped. This is accomplished using a *#Define OFFLINE* statement in the *opa.F90* module that chooses between the input of fields from a previous run or the time-integration of these fields, before the call to the tracer routine. This method has been demonstrated to work but is not yet ready for general use.

2.1.7 Particle tracking

The inherited NEMO-OPA model contained code for the evolution of floats seeded in the domain (*floats.F90*). The computation of the trajectories of passive particles in a multi-processor environment is complicated because particles can move from one tile to another. The inherited NEMO code worked correctly in this regard (although the 4th-order Runge-Kutta stepper scheme (*flo_4rk*) did not work for us) but the general structure of the module was not suited to our specific needs. As well, no turbulence scheme for the particle trajectories existed – something that is often important for studies of dispersion in shelf regions. Author Brickman modified the *floats* code to include a correct Random Displacement Model (Rodean, 1996) as the turbulence scheme, and simplified the I/O of particle positions to make it more suited for studies in coastal environments.

While the abovementioned modifications were tested to work correctly, because multiple particle tracking experiments are often required, online particle tracking does not prove to be practical due to the overhead of computing the dynamic fields. As a result, most particle tracking done by the authors using CANOPA fields, is done offline using output model fields and MatLab (or other) routines.

2.1.8 Other minor changes of note

A number of slight changes to the code and/or runtime environment were made. These included the addition of a *perpetual_forcing* variable that disabled calls that updated the

day counter. This has the effect of freezing all input forcing fields to the initial input fields thus allowing the model to equilibrate (spinup) to a specific, stationary forced state. Also added was a routine to output various shelf-specific quantities to a simple text file (module *sopa_mc.F90*), and the creation of a *run_info* output file to keep track of runtime information specific to CANOPA. A *run_params* input file was added to set CANOPA-specific variables.

2.1.9 Closing remarks

It should be appreciated that at the time that the authors inherited the NEMO-OPA code there were virtually no shelf applications of the model, and all specific code algorithms pertained to specific deep water applications, with little general utility. This fact, enacted through the often complicated *#Define* structuring of the code, often made reading of a given module (or *.F90* file) extremely difficult as the vast majority of the lines were dedicated to specific *#Define-keys* unrelated to shelf modelling. The authors eliminated as many of these *#Define* code-blocks as possible, at the expense of possible loss of compatibility of CANOPA. The code is also riddled with numerous modules with *relevant* code contained within complicated *#ifdef* – *#elseif* – $\cdots$ – *#endif* blocks whose readability would be improved by replacing the intervening code-blocks with *#include* files containing these code-blocks. The CANOPA code has accomplished this in a small number of cases.

The modifications to the code listed in this section were motivated by practical considerations regarding simplifying the work and runtime environments of CANOPA. Many of them were made with full awareness of the consequences with respect to NEMO code compatibility. As a result the authors have been hesitant to offer their code changes to the main NEMO consortium but, rather, have made their code (or a subset thereof) available to the local DFO community who have direct use for it. The future plan for CANOPA is to upgrade it by configuring the latest NEMO-OPA release on the MC domain, and then back-substituting the core CANOPA routines into this newer code. It is anticipated that this procedure will take about half a person year to achieve.

3 Model Setup

Simulations of a regional ocean model can be looked at as the model's response to boundary forcing, i.e. as an initial and boundary value problem. For realistic runs of a regional ocean model, interior TS fields are needed for initialization (and restoring). Also required are TS and velocity fields for lateral OBCs, and some form of surface forcing (usually derived from an atmospheric model).

3.1 TS fields

As mentioned above, monthly averaged 3D model TS fields were created using data from the BIO climate database. At startup, the model inputs 2 (bracketing) monthly fields and time-interpolates them based on the initial run date. To keep the model from drifting too far from climatology the NEMO-OPA code contains a simple restoring scheme whereby an adjustment is added (at each timestep) to the (complete) T (or S) fields:

$$dT \approx \frac{1}{\tau}(T - T_c)$$

where τ is a timescale, and $(T - T_c)$ is the difference between the model T and the climatology T_c, the latter time-interpolated from the monthly TS fields. CANOPA uses 2 timescales: one for the surface layer (*sdmp*, down to a depth *hdmp*), and one for the layer below (*bdmp*), with the 3 variables set in the namelist file. For spinup runs, where we want the model TS to be close to climatology (i.e. "diagnostic" simulations), timescales from 3-30 days are used in the top layer (usually <50m) and 30-90 days for the bottom layer. For long runs (multiple years) strong restoring to a climatology can be problematic, as this can dampen TS changes that one wants the model to simulate. In practice, we found that model drift in CANOPA was not a problem so that for long runs we used weak restoring with timescales of 900 days. (Note that other CANOPA users adopted much longer timescales with no drift problems.) An exception to the above (for the MC domain) is in the Gulf Stream region in the deep water offshelf where an option to use strong restoring ($\sim$ 30d) is activated in order to control drift in this region that is not important to our shelf solution but can affect it.

3.2 Open Boundary Fields

Open boundary TS fields are required and were taken as the values at the vertical boundaries of the monthly TS data files. Open boundary velocities were computed from the monthly TS fields using the thermal wind relation (with the understanding that these calculations do not include possible additional barotropic flows). Note that the model time-interpolates the above fields to the current model time.

The thermal wind results showed well defined shelf break velocity structures (jets) at the eastern and western boundaries, and noisy Gulf Stream-related features along the southern boundaries of the domain. The expected transport through SBI was not well defined. As well, no seasonal cycle of transport was evident. These observations held true using various assumed levels of no motion. Based on the above, it was decided to use the thermal wind velocities (with level of no motion at the bottom) to provide the vertical structure of the open boundary flows, and to add an analytic seasonally varying barotropic component at the SBI and SB valve locations (figure 2). These analytic functions entailed three parameters per location (mean, amplitude, phase), which were tuned based on comparing model results with interior data (see below).

3.3 Surface forcing

The NEMO-OPA code can accept various fo ms of surface forcing, at various timescales. CANOPA is typically run using daily values of atmospheric surface variables that it uses in bulk formulae to compute surface fluxes of mass and energy.

With the creation of various atmospheric r analysis programs, principally NCEP (National Center for Environmental Prediction) and ECMWF (European Centre for Medium-range Weather Forecasts), forcing data for extended periods of time have now become available to drive ocean models. The NCEP reanalysis period 1958-2000 has been analyzed to produce the Common Ocean-ice Reference Experiments (CORE) normal year (NY) of forcing (Large and Yeager, 2004). This 365 day "cyclical year" dataset is designed to represent the climatological atmospheric forcing with its climatological variability.

All the results reported herein used surface fo cing for the model derived from the CORE-NY dataset. Atmospheric wind velocity, surface temperature, precipitation (rain and snow), specific humidity and cloud coverage were rendered onto a daily timebase and spatially interpolated onto the MC grid. The CORE-NY fields were used to create a daily climatological circulation that contains the expected variability in the flow.

Other surface forcings adopted for model u se were the NCEP reanalysis fields (1958-2006), and CMC GEM 3-hourly output for the period 1998 to the present.

4 Model Validation I – General

The purpose of this section is to provide d tails of the model validation exercise and some general results. Validation is based on comparing model output from a CORE-NY simulation versus various climatological data. The next section provides more quantitative model-data comparisons. Useful model output can also be found in the companion report *Atlas of Model Currents and Variability in Maritime Canadian Waters* (Brickman and Drozdowski, 2012).

Generally speaking, model validation can be considered as an optimization problem in which tunable model parameters and inputs are adjusted to minimize model-data mismatches. A key question related to model validation is what controls are actually (and practically) available to tune the model results. For example, for the MC domain we expect that the open boundary flows directly influence the interior solution, as illustrated in figure 3. From this figure we can also deduce that the northward flow of warmer Newfoundland shelf water through Cabot Strait can influence the development of the GSL ice field, and that river inputs are also important. These inputs we have reasonable control over. On the other hand, the climatological circulation in the (e.g.) southern GSL can depend on the mean wind stress, and details of the bathymetry not resolved by the model – both of which in practice we do not have much control over. Related to the

above are the various parameters that can affect the model solution.

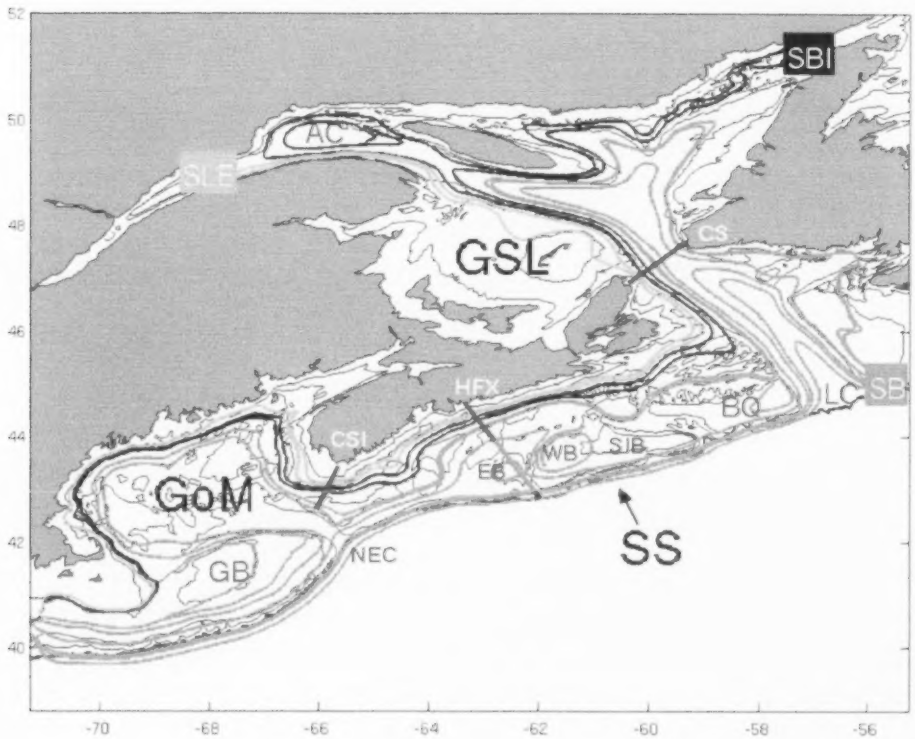


Figure 3: MC domain including streamlines illustrating the interior flow and its connection to the valve forcing. Although based on actual particle tracks, the streamlines should be taken as schematic. The 3 interior sections indicated are, from W-E: Cape Sable Island (CSI), Halifax (HFX), and Cabot Strait (CS). The Halifax section contains 3 subsections: coast-240m isobath (blue), coast-340m (blue + green), and coast-1000m (blue+green+black). Abbreviations are: GSL=Gulf of St. Lawrence, SS=Scotian Shelf, GoM=Gulf of Maine, GB=Georges Bank, NEC=Northeast Channel, EB=Emerald Bank, WB=Western Bank, SIB=Sable Island Bank, BQ=Banquereau Bank, LC=Laurentian Channel, AC=Anticosti, SLE=St. Lawrence Estuary, SBI=Strait of Belle Isle, SB=shelf break.

From the point of view of observations, the *detailed* climatological circulation in the MC region is not well known, with field programs sporadic in space and time. The result is that while some persistant features emerge, for example gyres around various banks, the fact that the measurements are sparse and non-synoptic in space and of limited time

duration means that interannual (and decadal) variability can dominate estimates of mean currents, or that a given "feature" is deduced from only one observation period.

The process of model validation when different types of validation data are available (i.e. velocity, ice, TS), is complicated by the fact that deciding what constitutes the best result can be subjective or application dependent. For example, a model that is "tuned" to both velocity and ice data may not be optimal for a particle tracking application which is best served by accurate velocity fields. Also note that models using a restoring term for TS are effectively assimilating these data, with the model-data mismatch controlled by the restoring timescale. This effectively biases any comparisons between the model's TS climatology and data.

Theoretically, uncertainties in model inputs and data do not limit an optimization procedure. However, in practice the number of parameters involved and the computation time it takes for a model run results in a numerical problem that cannot be practically solved. As well, no standards exist for accepting or rejecting a model based on validation performance, implicitly acknowledging that all models have strong and weak points and that the validation exercise should be considered illustrative at best. These considerations indicate that a more pragmatic approach to the validation exercise is warranted. With this in mind the approach taken was to attempt to simulate 2 robust metrics: the monthly mean transports through the Halifax and Cape Sable Island sections (Fig. 3), and the spatial-temporal development of the GSL ice field – i.e. system properties for which there is reasonable historic data and/or publications available. Also, as can be deduced from figure 3, these metrics are not locally determined but rather are related to domain-wide circulation properties. Thus they provide good overall measure of model performance.

To simulate these variables, the SBI and SB valve parameters were varied as these exert an obvious control on the interior solution. The ice model has a number of tunable parameters, but sensitivity runs showed that they had little effect on the development of the GSL ice field, so the default values (from the inherited ice namelist file) were used. The development of the ice field was found to be more sensitive to the air-sea heat flux coefficient, which was varied by a factor of 1.0-1.5. We note that modifying the temperature of the inflowing SB water was, surprisingly, found to have no noticeable effect on the development of the eastern part of the GSL ice field, especially compared to changing the heat flux coefficient. This is why the latter was chosen as a tunable parameter.

The approach to tuning these parameters mirrored a numerical method insofar as the effects of adjustments to parameters were monitored through comparison to data, but this comparison was qualitative in nature, governed by the necessity of achieving an endpoint in $O(< 20)$ iterations. We consider this approach to be commensurate with our goal of calibrating the model for ecosystem applications, and consistent with the fact that the validation data themselves are subject to considerable uncertainty. Aside from the valve parameters, the model has a number of other tunable parameters, related to

(for example) mixing and surface fluxes, which could affect the interior transports and ice simulation. Changes to these were not considered.

In order to produce a stationary starting field, the model is initially run for 30 days with perpetual January 1 surface and open boundary forcing. For this run, the model interior TS field is strongly restored toward climatology (timescales 3d for the surface layer, 30d elsewhere). With the valve forcing parameters and heat flux factor chosen, the CORE-NY run starts from the January 1 restart files. To allow for TS variability in the annual cycle runs, the restoring timescales are set to much longer values, with no vertical variation.

Because the model is not expected to initially be in balance with its surface and open boundary forcing, the model is run for a number of years and the domain-averaged T and S monitored. It was found that the timescale for equilibration varied with the restoring timescale, qualitatively levelling off for timescales > 500d, for which it took 2-3 years to achieve a reasonable repeat cycle (NB: domain-averaged T and S exhibit a seasonal cycle). For the validation run reported here, we chose a 900d restoring timescale and a 6 year run-length, and analyze output from year 6 of this run. Note that many flow properties were qualitatively similar from year-to-year, indicating robust model performance.

4.1 Validation Results

Model transport is validated for the Halifax section and the CSI section. Data for the former comes from Loder et al. (2003) and Anderson and Smith (1989). Data for the CSI section comes from Smith (1983). The space-time ice field from the model is compared to ice concentration data, created from ice charts (Drinkwater et al., 1999) and rendered onto the model domain. The original ice data spatial resolution (0.5° latitude X 1.0° longitude) is coarser than the model's so that only qualitative comparison is considered in this section. The timeseries of model total ice volume was compared to the climatological monthly ice volume derived from a 40 year monthly timeseries dataset from 1969-2008. A more quantitative comparison to the ice data is provided in the next section.

Fig. 4 shows daily timeseries of model transports versus monthly data for the inner Halifax section (coast-to-240m isobath = Hfx240) and the CSI section. The model solution exhibits considerable variability but captures the main features of the data, including the seasonal cycle of transport at Halifax, and the tendency of transport reversals at CSI. In the annual average, the model has a slight tendency to overestimate the transport (0.8 versus 0.7 at Halifax, 0.3 versus 0.1^+ at CSI). This tendency pertains to the Hfx340 and Hfx1000 comparisons as well (not shown). With respect to variability, the model solution compares favourably with data for which such estimates exist (see Fig. 4-b, cyan boxes, based on Anderson and Smith, 1989).

The development of the ice concentration field for early January, February, and March, is shown in Fig. 5. Taking into consideration the coarser spatial resolution of the ice

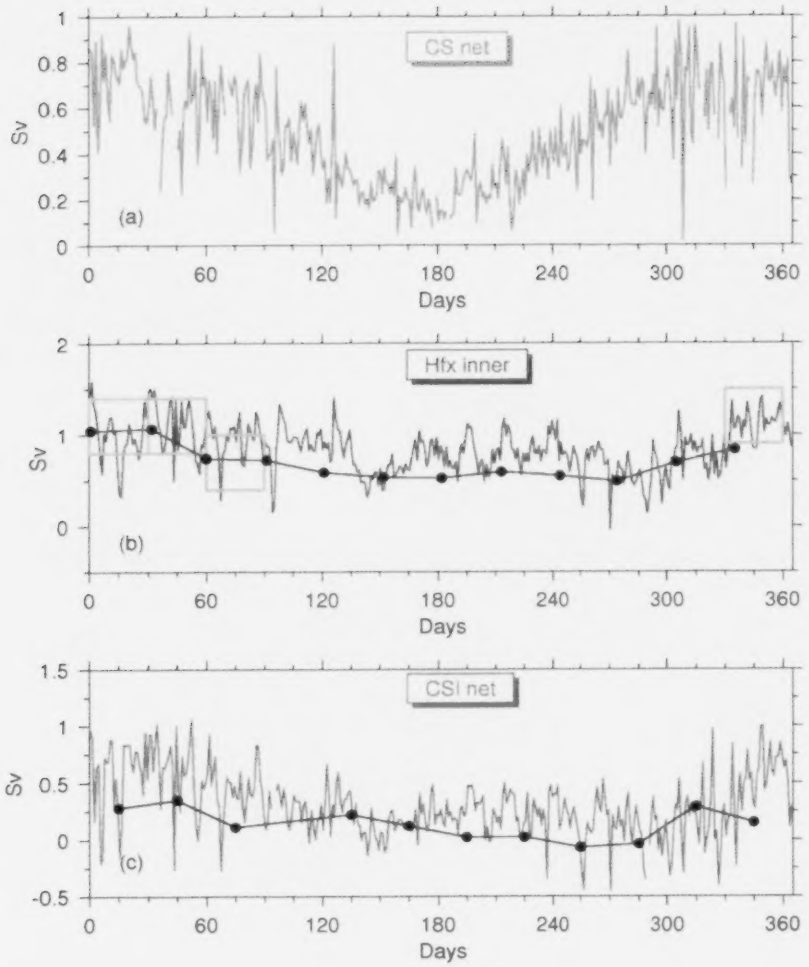
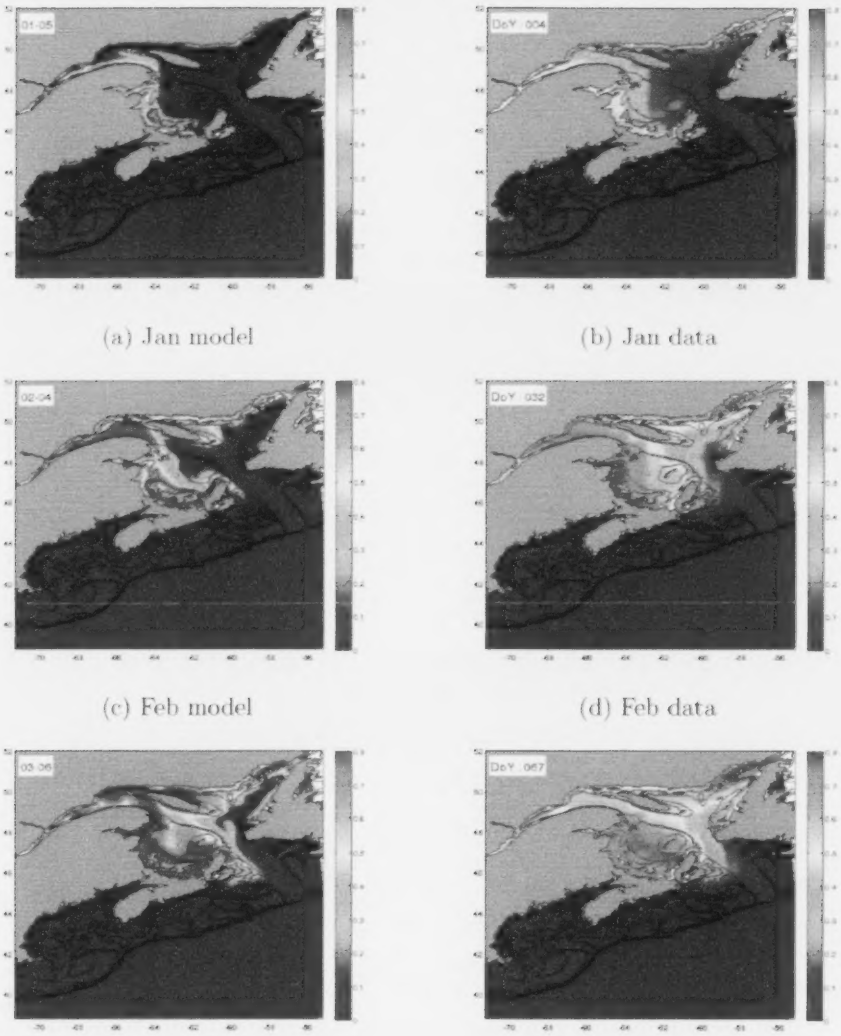


Figure 4: Model transports versus data. (a) Model's net transport through Cabot Strait (shown for illustrative purposes). (b) Halifax section transport – coast to 240m isobath. Black line is Loder et al.(2003) estimate. Cyan boxes bound the estimates from Anderson and Smith (1989). (c) CSI transport. Black line is data from Smith (1983).

data, the model is seen to capture the main features of the ice field, including the lower ice concentration in the eastern GSL due to the inflow of warmer Newfoundland waters. The model predicts an earlier final retreat of the ice field than the data (in April, not shown).

This latter tendency is reflected in the ice volume comparison (Fig. 6) which shows an excellent agreement with the data for the ice growth months (within 10% of the mean), with a larger error for April (although still within one std of the data).



(a) Jan model

(b) Jan data

(c) Feb model

(d) Feb data

(e) Mar model

(f) Mar data

Figure 5: Development of ice concentration field. Left column is model, right is data. Scales are equivalent.

The tendency of overestimating the interior transports while underestimating the retreating ice volume can be related to the tuning of the valve parameters, in particular the SB inflow. In general, decreasing the shelfbreak transport decreased the transports through the Halifax and CSI sections, improving the model fit. However, this also affected the inflow of warm water through Cabot Strait into the GSL, which increased the ice volume and produced less open water in the southeast Gulf, generally degrading

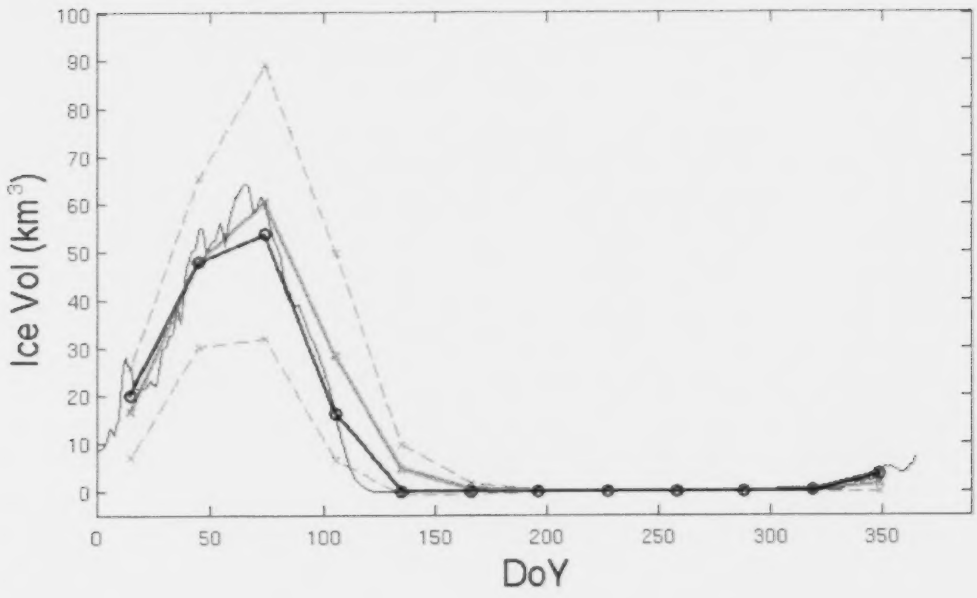


Figure 6: Comparison of model ice volume versus data. Red lines are monthly data, mean ± 1 std. Black lines are model: thick = monthly mean; thin = daily values.

the model-data ice comparison. This illustrates the subtle interplay between the open boundary conditions and the interior solution.

The final valve parameters (mean, amplitude, phase (in days from Jan. 1)) were 0.4Sv, 0.3Sv and 30d for SBI, and 3.0Sv, 0.5Sv and -60d for SB. Overall, we consider that the model does a good job of reproducing the transport and ice data using these open boundary conditions. However, it is clearly possible that other changes could have produced a better result but it was beyond the scope of this study to pursue this.

5 Model Validation - II: Quantitative comparison to current meter and ice data

This section provides a more comprehensive quantitative comparison of model output to current meter (CM) and ice data. As mentioned above, current meter measurements are sparse in space and of limited time duration so that interannual (and decadal) variability can dominate estimates of mean currents, and a given circulation "feature" may be deduced from only one observation period. This is problematic when considering comparing climatological circulation model output with data as the latter is potentially poorly defined. The first part of this section looks in detail at the space and time distribution

of current meter data in Maritime Canadian waters.

The model validation starts with a comparison of model currents versus data on a vector-by-vector basis – to look at the distribution of model errors in a domain-wide or "global" sense. This is followed by an analysis of model errors on seasonal timescales. In this case, the data are grouped in space and time in order to create a climatology of current meter data to which the model climatology can be compared.

The section ends with a comparison of the model's ice climatology to ice data. Ice variables computed are the day of first ice appearance, day of last ice appearance, and peak ice volume. The ice data span the years 1963-2011, and we compare the model's climatology to timeseries of these variables to the get a sense for how the model does with respect to the mean and variable data.

5.1 Current meter data distribution

The current meter data were extracted from the ODI database (Gregory, 2004) for the the period 1960-2010 on a region covering the east coast of Canada. The data exist as monthly means or submonthly means (if the series only covered a part of a month), although individual series are available by special request. CM deployments peaked in the late 1970s to early 1980s, with the majority of deployments ($\sim$ 3000) in the MC domain region. Deployment length has increased from an average of 60 days in the 1970s, to 200 days presently, due to technological developments. Associated with this is increased use of acoustic Doppler current meters (ADCPs) which provide much higher vertical resolution. (See Gregory, 2004, for further details.)

The CANOPA Normal Year run is a simulation of the climatological circulation in the region. The objective of this subsection is to provide an idea of the degree to which a circulation climatology can be obtained from the CM data, and to help with selection criteria for the data. To do so, we look at the (space/time) distribution of the data grouped in 0.2x0.2 degree cells in the horizontal and a 0-250m vertical depth range. The choice of horizontal binning was found to adequately preserve the basic spatial distribution of the data, and group together profiles that were closely spaced. The vertical depth range spans from surface to bottom for about 85% of the shelf region, thus capturing all the data for most profiles as is desired for an overview analysis of this type.

Summaries of the total monthly coverage is shown in figure 7 (spatial plot of color coded symbols of the number of months of data) and figure 8 (histogram of the data). The spatial distribution of locations indicates (not surprisingly) the difficulty in creating a climatology from CM data, with little coverage in the central GoM, through most of the Laurentian Channel, and northeastern SS. Most locations in the southern GSL and western SLE, and to a lessor extent the central SS, contain ≤ 5 months of data. Overall, about 57% of locations have ≤ 5 months of data, with 14% having > 15 months (figure 8). Note that this analysis does not distinguish between months so, for example, a location

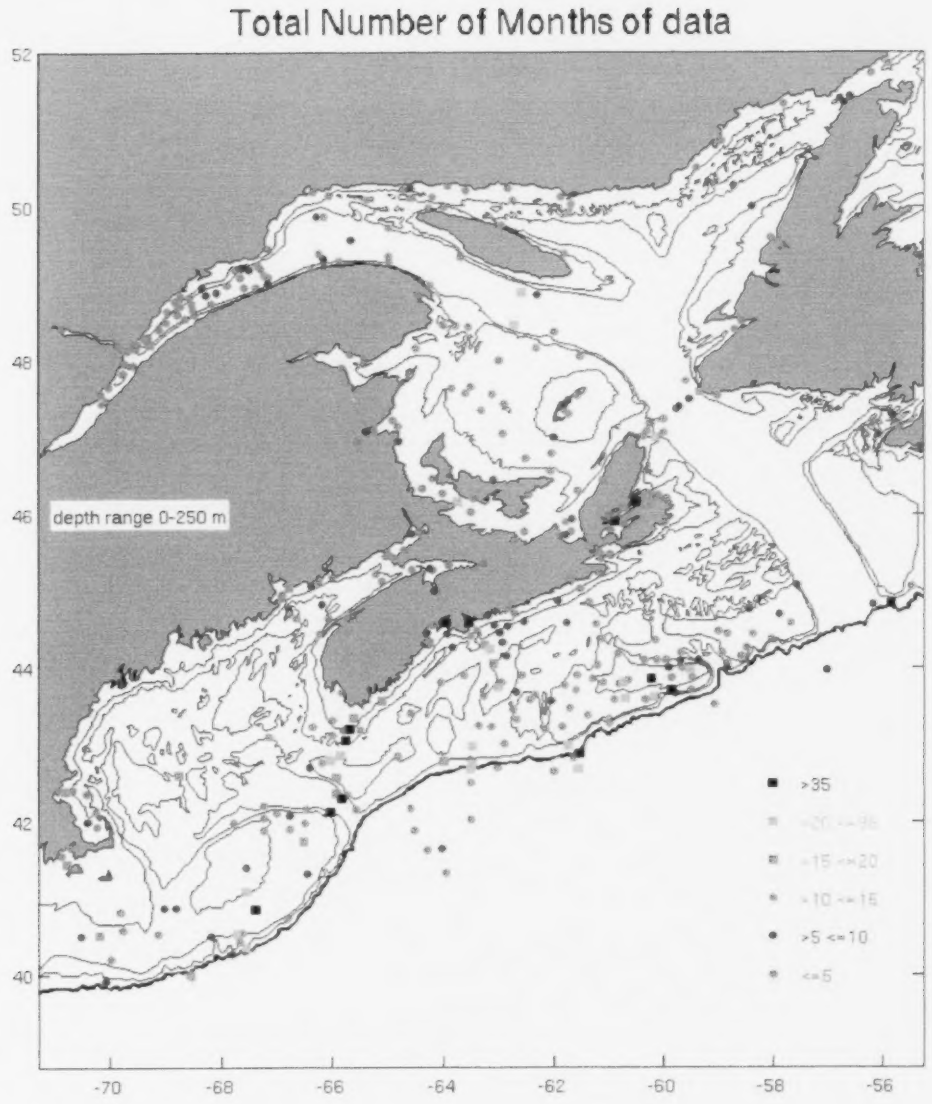


Figure 7: Colour coded symbol plot of the total number of months of data in the 0-250m depth bin.

with 5 months of data could have 5 years of the same month being sampled.

Figure 9 shows the number of distinct years of data spatially, while figure 10 is a histogram

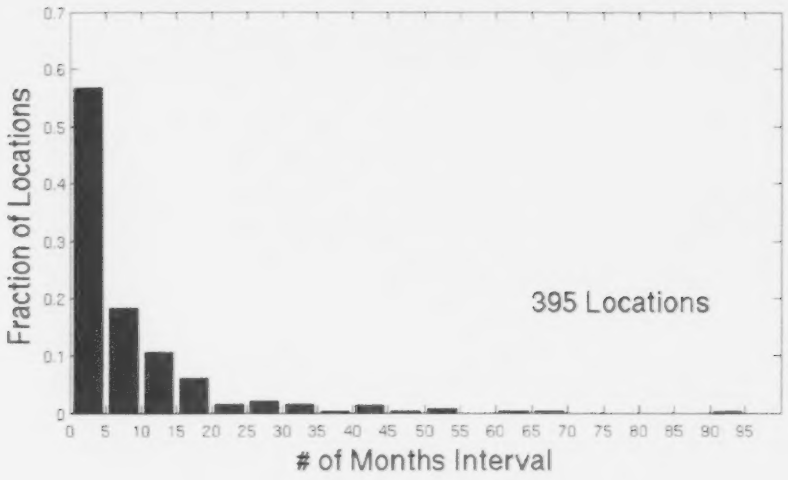


Figure 8: Histogram of the number of months of data in the 0-250m depth bin.

of the data. The analysis shows that the majority of locations have data from only 1 year (51%, figure 10, open red circles figure 9). Notably, the southern GSL is mostly populated by single distinct years of data. Thus any climatology (monthly, seasonal, annual) for these locations is subject to undersampling effects that could result in misrepresentation of the true mean circulation. Note that only 22% of locations have 3 or more distinct years of data (figure 10), so that if this were a selection requirement we would lose about 78% of the data.

With respect to constructing an annual mean velocity dataset, if we require at least 7 (9) distinct months of data then from figure 11-btm we find that only 30% (20%) of locations satisfy this requirement.

This analysis illustrates the potential problems associated with creating a climatology from CM data, with the majority of the shelf area with little or no coverage and the majority of sample sites having a single occupation only. This complicates any comparison between a circulation model climatology and CM data as it is difficult to determine whether model-data discrepancies reflect poor model performance or anomalous data. For example, if we consider the flow in the surface layer or in shallow water, which is strongly influenced by the wind stress, a comparison between the model's data and monthly data derived from a single year (i.e. sample size of 1) could simply reflect the difference between that year's wind stress and the model's forcing climatology.

Ideally we would like to limit our comparisons to locations with repeat sampling, but as shown this strongly reduces the available sample size. Therefore compromises have to be made regarding data selection criteria. For example, when considering goodness-of-fit metrics averaged over the complete domain, or subregions, a choice exists between having

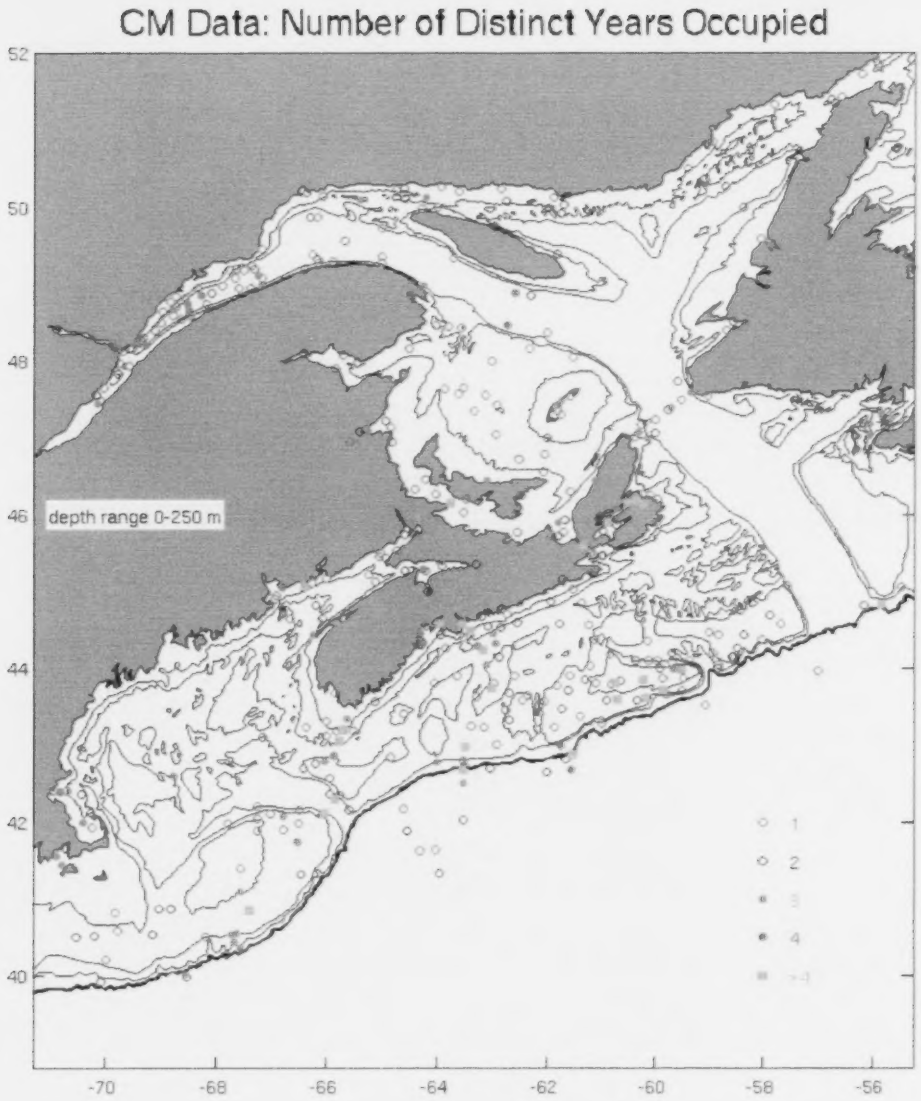


Figure 9: Spatial symbol plot of the number of distinct years of data in the 0-250m depth bin.

strong selection criteria that produce a small sample size of higher confidence data, or weak selection criteria that result in a larger sample size of lower confidence data. In the comparison between model output and CM data on seasonal timescales we will provide

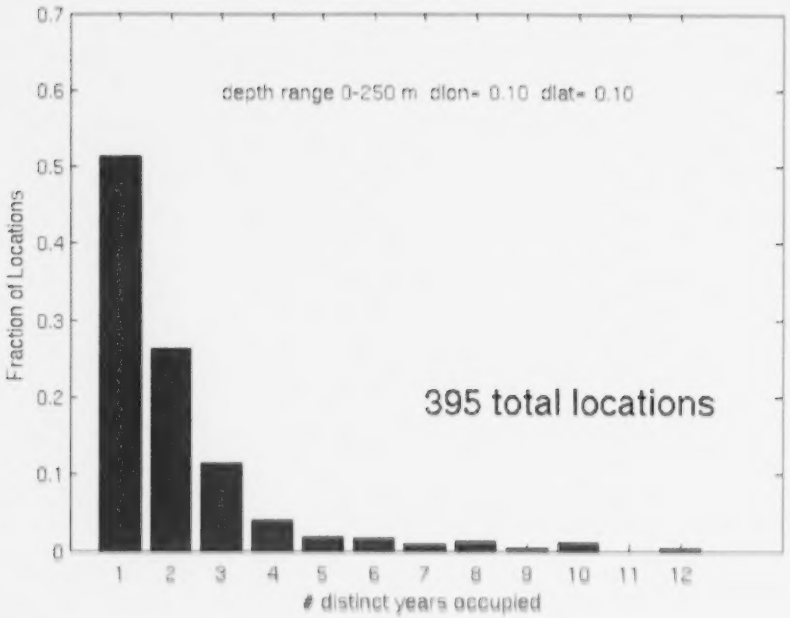


Figure 10: Histogram of the number of distinct years of data in the 0-250m depth bin. (NB: dlon, dlat refer to half grid cell dimension.)

an example of model skill for two data selection criteria.

5.2 Velocity Comparison: Global Analysis

The Normal Year model currents were validated using current meter data from the ODI database. For each observation at x, y, z spanning yeardays t_i to t_f, CANOPA Normal Year currents were computed at the corresponding position, averaged from t_i to t_f. Because we are comparing individual observations against modeled climatological currents, it is expected that there would be a large variability in the model–data agreement. Regions with high interannual variability are expected to compare poorly with model results, as opposed to regions of low variability where the mean, as represented by the model climatology, should be closer to individual observations. Recall from the previous subsection that data temporal density is such that distinguishing high versus low variability locations is rarely possible. Nevertheless, due to the large sample size of data ($\approx$13k (x, y, z, t) records), we expect a central tendency in the distribution of errors. Figure 13 is a scatterplot of model speed versus observation speed. As expected, the cloud of points is roughly centered over the model equals data line, even though there is about an order of magnitude scatter in the distribution. The relationship is investigated more closely in Figure 14 which shows a histogram of the difference between modeled and observed

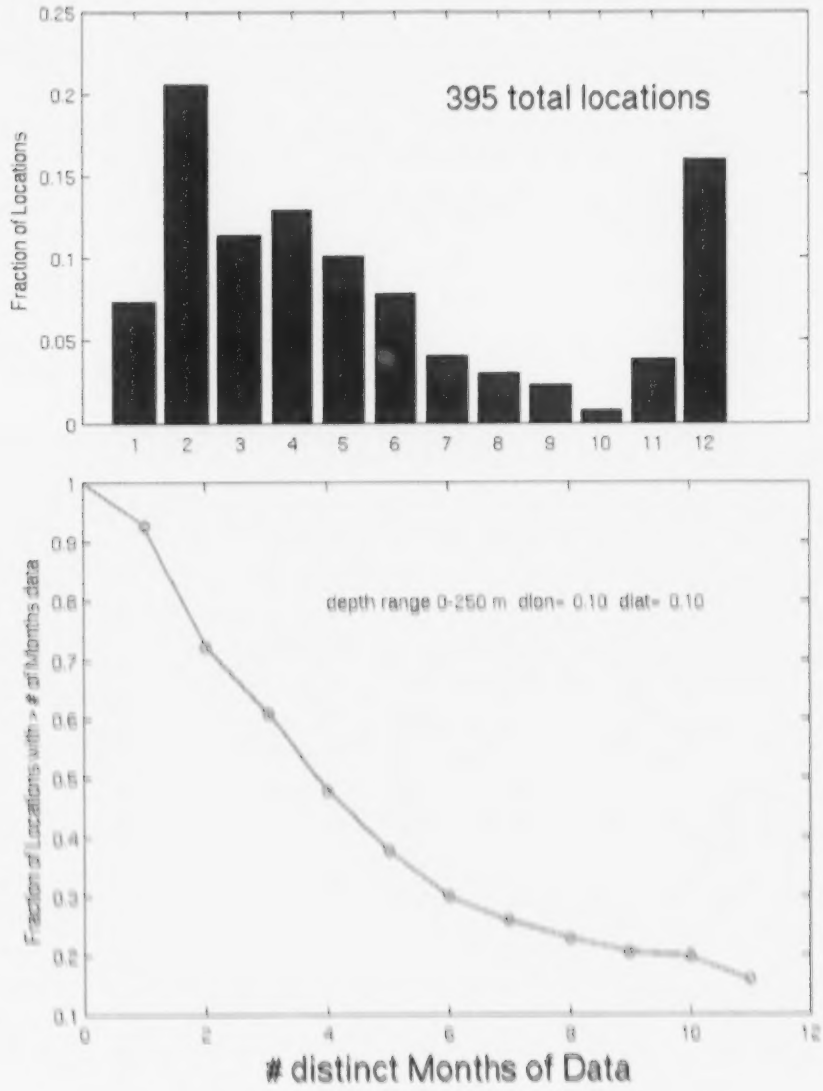

395 total locations
Fraction of Locations
depth range 0-250 m dlon= 0.10 dlat= 0.10
Fraction of Locations with > # of Months data
distinct Months of Data

Figure 11: Top: Histogram of the fraction of locations that have N distinct months of data (for the 0-250m depth bin). Btm: Fraction of locations with $>$ N distinct months of data. (NB: dlon, dlat refer to half grid cell dimension.)

values. The distribution is Gaussian-like with a median of 0.003 m/s and a mean of 0.0205 m/s. 90% of the errors are in the range -0.15 to 0.27 m/s. The distribution is slightly skewed towards larger model values (i.e. the mean is to the right of the median).

Figure 15 shows the distribution of the ratio of modelled to observed speed. As this variable is bounded by zero and infinity a log scale is appropriate. For this metric we find an almost log-normal distribution with approximately 90% of model to data ratios within an order of magnitude.

To this point we have only considered magnitudinal differences and ignored the directional component of current. Figure 16 shows the distribution of the magnitude of the modelled-

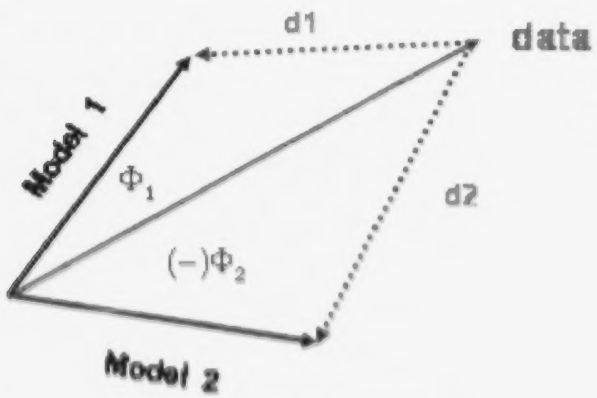


Figure 12: Illustration of vector differences. The data vector is in red. d1 and d2 are model–data difference vectors. The angles ϕ_1 and ϕ_2 are angle errors with sign defined by the right hand rule: data $\times$ model

minus-observed vector difference (V_d') and the angle error (ϕ) (as defined in figure 12). If the model was perfect (i.e. zero mean) with random errors then the difference magnitudes ($\in [0 \rightarrow \infty)$) could be expected to follow an exponential-like distribution, while the angle errors would be expected to follow a Gaussian distribution. The difference magnitudes (top panel) do exhibit an exponential-like decay with magnitude but do not have a maximum at zero magnitude. Rather, the bin with the biggest count is $\approx 2cm/s$. The angle error distribution has a mean of ~ -4 deg, but does not decay towards zero for larger values. Instead it stays about constant for angles beyond ± 50 deg (with a slight increase towards ± 180 deg), resembling a combination of uniform and Gaussian distributions. This would suggest that there is a subgroup of currents where we are doing poorly, with the comparison resembling white noise. Attempts were made to isolate this badly behaved group based on region, depth, topography and speed. For example, it is a reasonable hypothesis that the model's angle error distribution would be uniform when current magnitudes are small. However, neither this nor other factors seemed connected with this white noise region. We suggest that this behaviour of angle errors is an expected result of comparing climatologicial model currents to current meter data on a vector-by-vector basis, i.e. the result of comparing a model's estimate of the mean.

at a given location and time, to a data estimate (or estimates) that can be considered to be drawn from a distribution with unknown mean and variance. (Think of each datum being a random sample from a "current rose" diagram, which is compared to the model's estimate of the mean for that location and time.)

One more comparison is included to summarize this subsection. So far the comparison restricted itself to data-model pairs from corresponding time and location. However because we have large data set of currents, the speed distribution can be considered to represent a climatology for this region, which can be compared to the model's version. Figure 17 shows that the model does a good job in reproducing the domain-wide distribution of current speeds

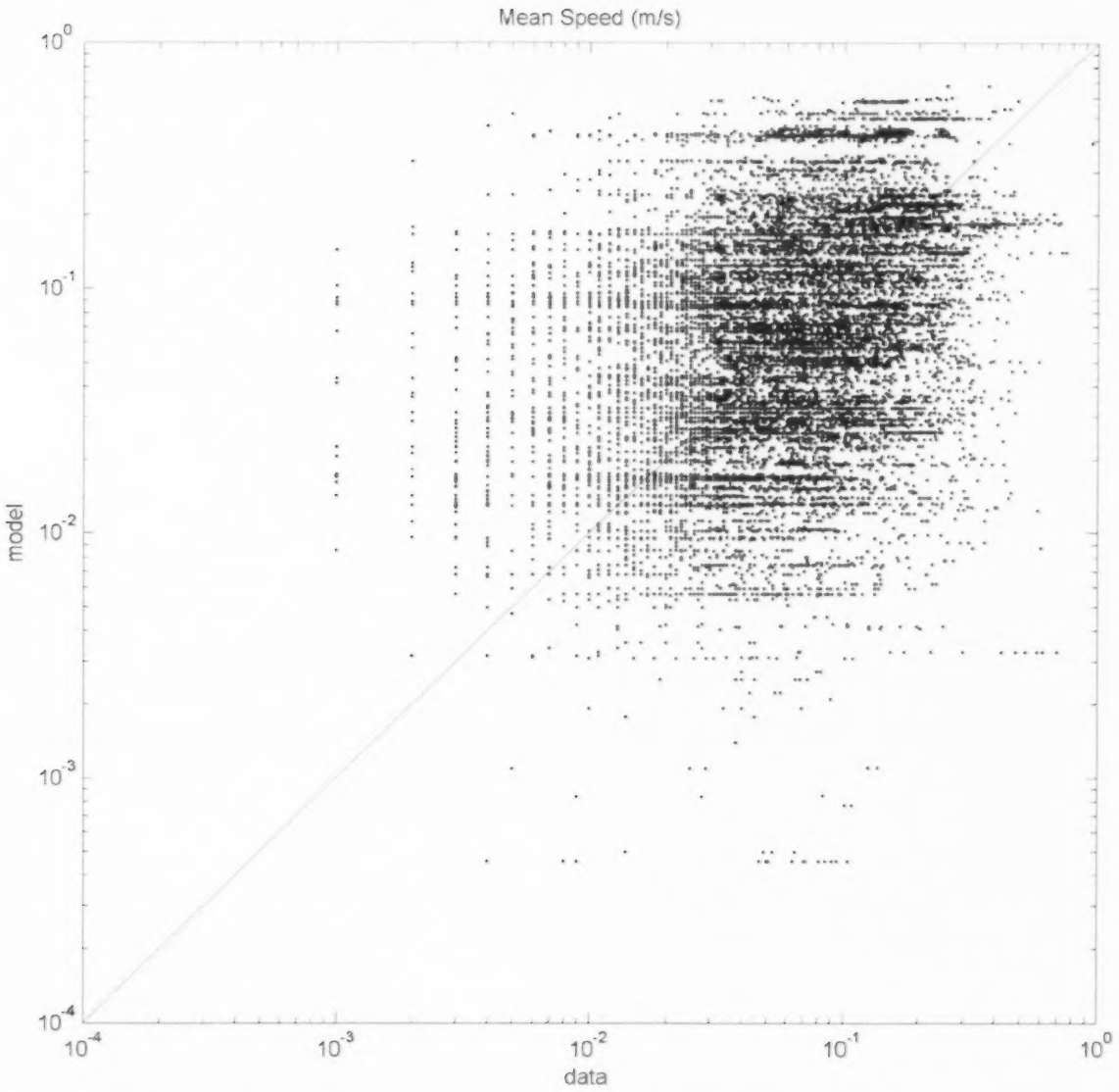


Figure 13: Scatterplot of model speed versus observation speed (m/s). Note the log scale. The model equals data line is drawn through the data.

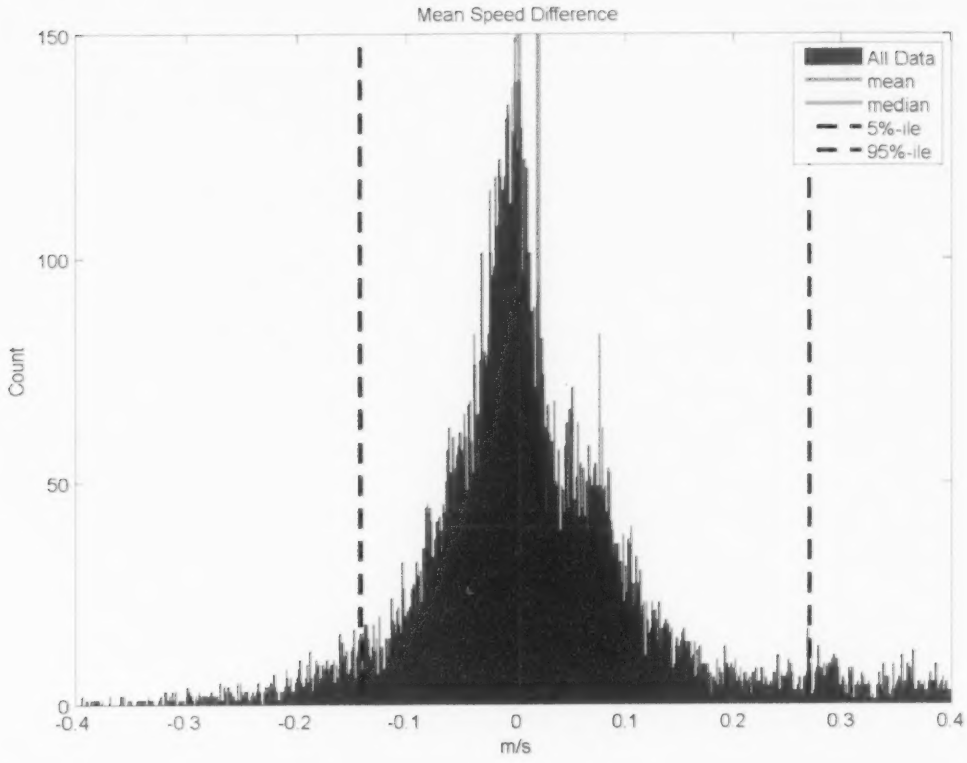


Figure 14: Frequency distribution of model–observation differences for monthly mean speeds.

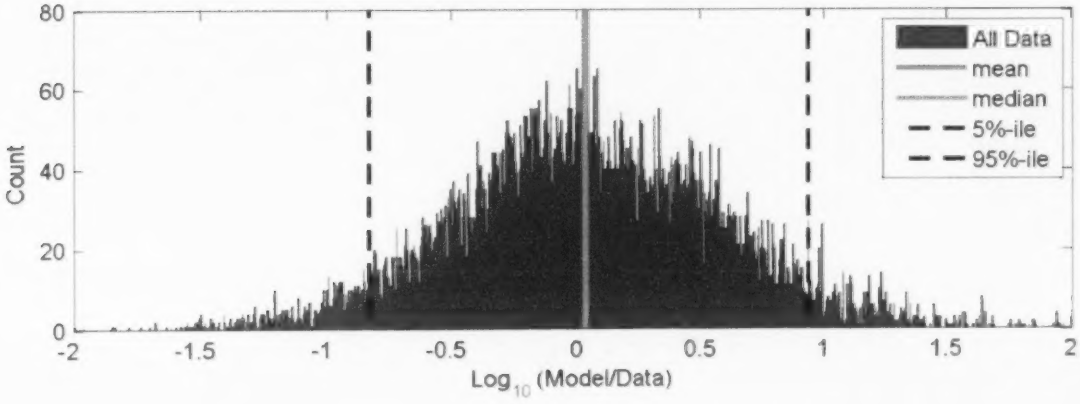


Figure 15: Frequency distribution of the ratio of model to observed speed on a log scale.

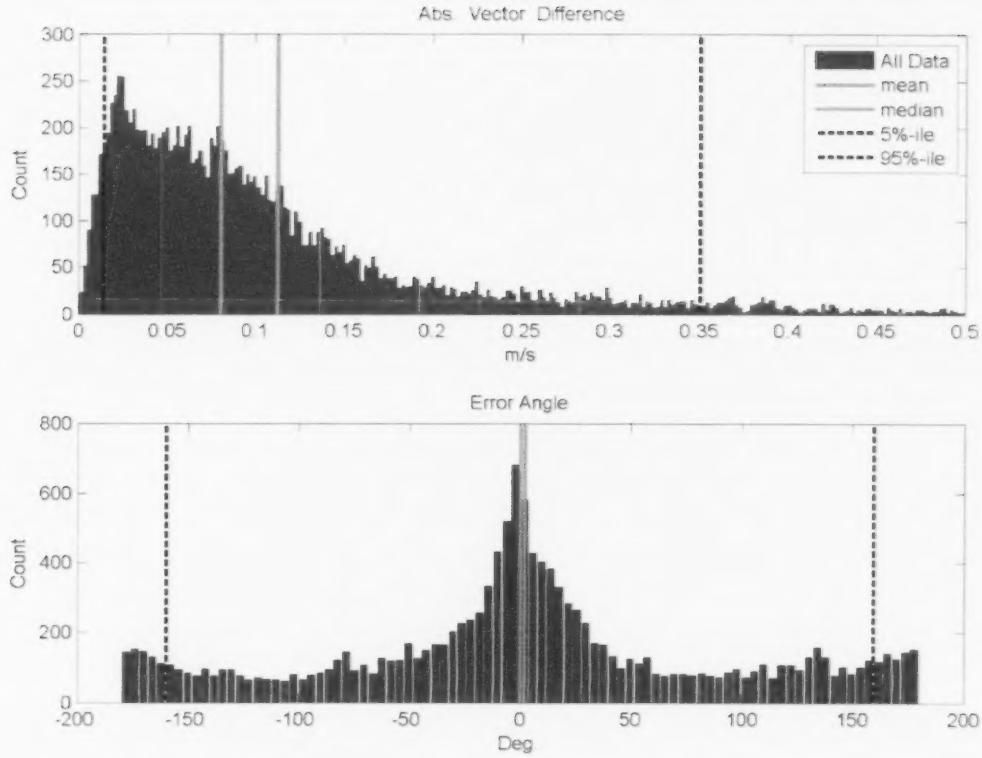


Figure 16: Magnitude of model-observation vector differences (top) and error angle (bottom)

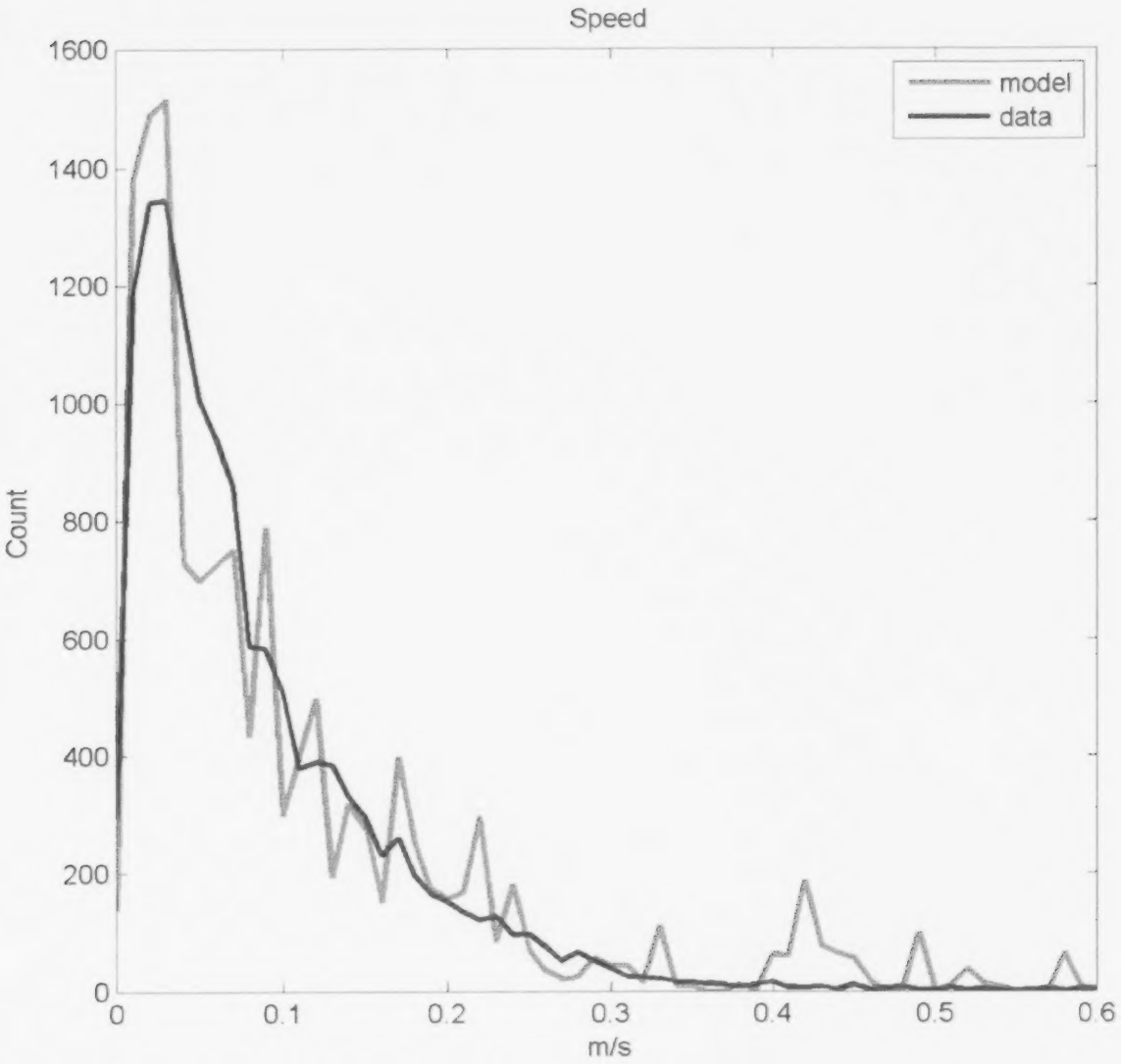


Figure 17: The distribution of speeds for both model and observation.

5.3 Velocity Comparison: Seasonal Climatology

Seasonal estimates of currents are only possible in regions of high data clustering. In order to create a seasonal climatology, the data were grouped into cylinders of radii up to a few tens of kilometers in the horizontal (figure 18), with variable thickness bins in the vertical (0-20, 20-50, 50-100, 100-200, 200-500, 500-1000, 1000-2000, >2000m). The cylinders were hand picked from areas of high data density, resulting in 205 distinct cylinders with 1562 possible data. Note that this differs from the 0.2°x0.2° squares used in the data distribution analysis above.

Data selection criteria were applied based on the number of observations per season per cylinder-bin (Nsam), the number of unique years per season per cylinder-bin (Nu), and the signal to noise ratio (SNR). The latter was defined as the ratio of the magnitude of the mean current in each cylinder-bin divided by the standard error (SE) defined as

$$\mathrm{SE} = \frac{\sqrt{\overline{\|U - \bar{U}\|^2}}}{\sqrt{N}}$$

where N is the number of data. Strong and weak selection criteria were used. Weak selection (following Hannah et al. 2001) had Nsam > 1, SNR >0.5, but did not require more than one unique year of data (Nu=1). Strong selection required more than one unique year of data (Nu> 1), more than one month of data per season, and a SNR >0.5 (Nu > 1, Nsam > 1, SNR >0.5). Note that in this case Nsam > 1 does not necessarily mean distinct months of data in a given season so that the same month sampled in different years is acceptable under these criteria.

The mean seasonal current was computed inside each cylinder, averaged over each depth bin, using data available after the rejection criteria were applied. No attempt was made to weight the average in any way, it is simply the mean of the monthly (or greater than 14 day) means. The count in each bin varied from 1 to a few tens of observations. Note however that even a count of 1 can be meaningful since it is a comparison based on sufficient data to pass the rejection criteria.

Model currents for the comparison were taken from the seasonal mean field and nearest grid point to the cylinder centers. Due to nearness to the open boundaries, the Strait of Belle Isle, Laurentian Channel and Western Gulf of Maine regions were excluded from the comparison. For analysis purposes, the Maritime Canada region was further subdivided into blocks (figure 18). The statistic used in the comparision is the overall model skill – the ratio of vector difference to observed kinetic energies (r from Hannah et al. 2001).

$$r = \sum \|V_d\|^2 / \sum \|V_o\|^2 \qquad (2)$$

Figure 19 shows the skill results for the weak criteria case in table form, while figure 20 provides the number of data associated with each skill table entry. Following Hannah et al. 2001, regions where r<=0.5, good agreement, was color coded green, regions of fair

agreement 0.5<r<=1.0, was color coded yellow. All other data, r>1.0, was left white. Note that in these tables the "All"/"Annual" rows/columns are computed by regrouping the data and re-calculating r. Because r is a non-linear statistic these values are not the same as weighted averages calculated using the count data in figure 20 (although such differences proved to be small).

The overall skill (all regions, seasons and layers) is 1.05, just slightly outside fair but still quite good considering this is a large region, the entire water column, and places where data might be too sparse to form a reliable climatology. The weak selection criteria resulted in the loss of about 20% of the possible data (selected 1242 out of 1562). If we look at the layers (for all regions and season), there is fair skill for 0-50m and 500-2000m with best agreement below 2000m. Looking at seasonal differences for all the regions combined, we see that fall is the season with best skill with only the 100-500 lacking good or fair skill.

Looking a different regions, we see highest skill for the Cabot and Cape Sable regions. For those 2 regions, most of the boxes are green or yellow. The Cabot region skill is lowest in the winter, and about the same from spring to fall. The Cape Sable result is best in the winter which is consistent with the Hannah et al. 2001 validation. The worst agreement is in Scotian Slope region. There are interesting seasonal variations in the regions. For example, in the Anticosti region we get good agreement in the summer but poor during the rest of the year. Other regions (St. Lawrence, Southern Gulf, and Sable Island) the model does well in 2 of the 4 seasons. For the Halifax region the agreement is fair for all seasons except summer (note the $r = 131.04$ in summer. This is based on a single observation which disagrees wildly with the model; perhaps related to NS current reversal).

Another interesting observation is that in the Scotian Slope region, the skill above 500m is always poor but quite good for the deeper layers. This suggests possible affects of the Gulf Stream intrusions and/or shelf break current which would be more pronounced in the top layers. These types of intrusions are not climatological and we have no hope of modelling these with the Normal Year forcing.

The above results used the weak data selection criteria. Use of the strong criteria eliminated many cylinder-bins from the tables, and reduced the total number of data from 1242 to 718 – a loss of 54% of the possible data. The overall skill (all regions, seasons and layers) was slightly degraded (1.08 vs 1.05), but comparisons in certain regions (Halifax, Cabot Strait and Southern GSL) benefited from the stronger rejection criteria (figure 21). For Cabot and Halifax the overall skill improved from fair to good, while for the southern GSL the agreement improved from poor to fair. In general the increased skill can be attributed to the elimination of bins with poor agreement due to insufficient data based on the strong criteria.

An interesting pattern emerges if we look at the data in figures 19 and 20 in terms of model skill as a function of the number of data. If we consider each bin in the 8 regions as

a sample (ignoring the "All" row and "Annual" column, i.e. the summary calculations) and scatterplot the model skill versus #-data per sample (not shown) we find that the mean skill and standard deviation decrease abruptly at about 13 data points per sample. For <13 data points the mean skill and standard deviation are 3.75, 12.59, while for higher data counts the values are 1.04, 0.57. (Recall that because r is a non-linear statistic these values are not the same as grouping all the data in the two sample sets and computing r, as was done in the "All" and "Annual" values in the skill table.) We explain this behaviour in the following way: Consider that the model has some true skill that we are trying to determine. Each sample defined above represents a depth bin in a particular region for a particular season. For each sample, the model provides estimates of the velocity at M model grid points with some distribution of skill, independent of the number of data in the sample (N). If N is small (evidently $<\sim$13) then estimates of the model skill are highly variable and tend to underestimate the skill (r high), while for larger N, estimates of r are less variable and likely better approximate the true model skill.

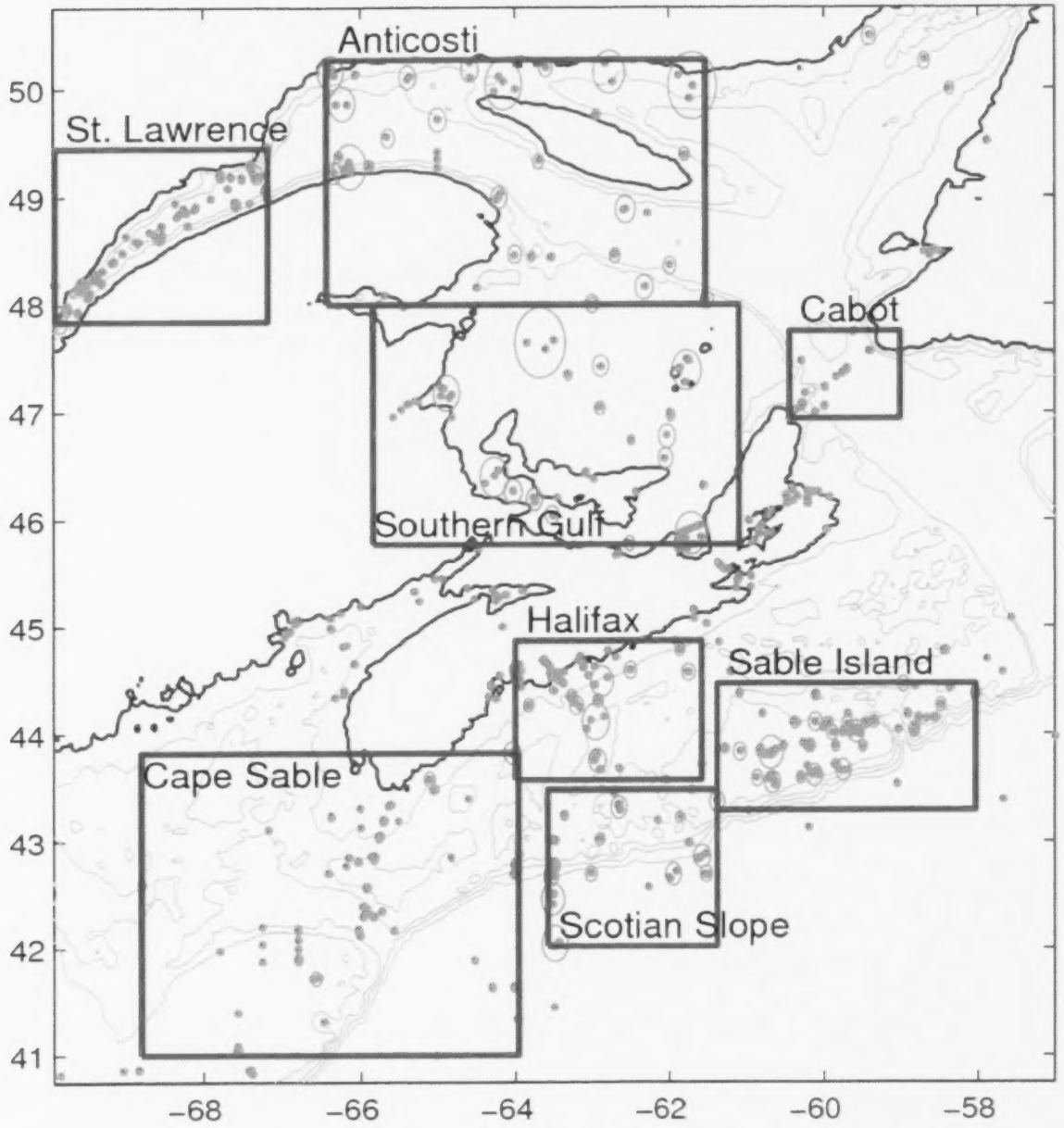


Figure 18: Horizontal distribution of data. Boxes define regions used for the analysis

ALL Regions

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	1.05	1.20	1.00	1.13	0.85
0-20	0.79	1.29	0.70	0.70	0.73
20-50	0.99	1.05	1.10	0.99	0.84
50-100	1.38	1.21	1.47	1.81	0.99
100-200	1.47	1.72	1.31	1.75	1.05
200-500	2.52	1.53	1.57	7.23	1.11
500-1000	0.58	0.44	1.86	1.63	0.28
1000-2000	0.63	0.56	1.25	0.70	0.34
>2000	0.52	0.66	0.02		

St. Lawrence

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	1.11	1.43	0.96	1.25	0.95
0-20	0.69		0.64	0.82	0.94
20-50	1.48	1.44	1.65	1.38	
50-100	1.62	1.33	1.81	1.52	
100-200	1.72		2.10	1.42	
200-500	3.47	1.66	5.96	1.04	1.99

Anticosti

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	0.76	1.49	1.64	0.66	1.28
0-20	1.17	0.77	1.74	1.67	
20-50	0.69	2.49	1.54	0.62	1.19
50-100	0.86	1.78	1.53	0.75	0.68
100-200	1.18		0.56	1.24	
200-500	3.34	3.08	5.38	1.44	6.94

Southern Gulf

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	1.17	1.80	0.93	1.86	0.81
0-20	1.10	1.69	0.93	2.03	0.76
20-50	1.61	1.94	0.93	0.46	3.00
50-100	3.18	3.23		2.67	

Cabot

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	0.53	0.85	0.41	0.45	0.52
0-20	0.39		0.37	0.33	0.50
20-50	0.64	0.80	0.41	0.70	0.57
50-100	0.46	0.82	0.50	0.83	0.19
100-200	1.17	1.32	0.76	1.14	1.20
200-500	0.57	0.55	0.61	0.75	0.42

Cape Sable

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	0.86	0.62	0.80	1.00	0.94
0-20	0.71	0.62	0.91	0.76	0.67
20-50	0.73	0.48	0.67	0.76	0.98
50-100	0.96	0.51	0.79	1.28	1.04
100-200	0.79	0.67	0.77	0.79	0.94
200-500	3.38	1.60	2.03	14.23	1.30
500-1000	0.18	0.14			0.25
1000-2000	1.01	0.47	0.92	0.90	1.89

Halifax

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	0.69	0.61	0.80	2.54	0.56
0-20	0.70	0.73	0.68	131.04	0.42
20-50	0.79	0.62	0.90	1.38	0.81
50-100	0.47	0.33	0.61	1.33	0.41
100-200	1.10	1.04	1.27		1.03
200-500	4.30	2.65	12.49	2.50	8.71

Scotian Slope

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	4.94	4.11	3.31	11.26	3.52
0-20	3.74	4.58	6.89	2.13	3.63
20-50	5.18	2.65	3.25	32.30	5.23
50-100	8.35	6.06	4.81	24.72	7.66
100-200	6.31	17.78	2.27	23.27	4.42
200-500	2.99	2.51	1.04	7.72	0.86
500-1000	0.58	0.46	1.86	1.63	0.28
1000-2000	0.59	0.57	1.27	0.65	0.34
>2000	0.52	0.66	0.02		

Sable Island

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	1.23	2.07	0.78	1.33	0.80
0-20	1.44	3.37	0.76	0.97	0.83
20-50	0.92	1.22	0.41	1.48	0.62
50-100	1.65	2.19	1.40	1.59	1.29
100-200	0.88	1.09	0.84		0.84
200-500	0.95	0.88	0.96	1.00	0.97

Figure 19: Model Skill for various regions and layers – weak selection criteria.

ALL Regions

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	1242	280	319	362	281
0-20	258	50	71	77	60
20-50	358	82	92	101	83
50-100	303	68	75	96	64
100-200	155	36	40	44	35
200-500	114	29	28	32	25
500-1000	25	7	5	6	7
1000-2000	27	7	7	6	7
>2000	2	1	1	0	0

St. Lawrence

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	146	9	58	70	9
0-20	46	0	20	20	6
20-50	38	3	14	21	0
50-100	28	2	12	14	0
100-200	19	0	8	11	0
200-500	15	4	4	4	3

Anticosti

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	93	16	21	43	13
0-20	12	1	6	5	0
20-50	32	6	6	14	6
50-100	32	7	6	14	5
100-200	8	0	1	7	0
200-500	9	2	2	3	2

Southern Gulf

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	32	9	6	11	6
0-20	20	4	4	7	5
20-50	8	2	2	3	1
50-100	4	3	0	1	0

Cabot

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	79	17	16	22	24
0-20	10	0	3	4	3
20-50	24	7	3	5	9
50-100	16	3	4	6	3
100-200	15	4	2	3	6
200-500	14	3	4	4	3

Cape Sable

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	315	74	70	91	80
0-20	54	13	12	16	13
20-50	82	19	19	21	23
50-100	78	16	15	27	20
100-200	63	15	15	17	16
200-500	32	9	8	9	6
500-1000	2	1	0	0	1
1000-2000	4	1	1	1	1

Halifax

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	103	33	24	11	35
0-20	22	9	4	1	8
20-50	28	7	8	4	9
50-100	30	9	7	4	10
100-200	15	6	3	0	6
200-500	8	2	2	2	2

Scotian Slope

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	195	50	50	49	46
0-20	12	3	3	2	4
20-50	39	10	12	9	8
50-100	41	9	9	12	11
100-200	23	7	7	6	3
200-500	32	8	7	9	8
500-1000	23	6	5	6	6
1000-2000	23	6	6	5	6
>2000	2	1	1	0	0

Sable Island

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	279	72	74	65	68
0-20	82	20	19	22	21
20-50	107	28	28	24	27
50-100	74	19	22	18	15
100-200	12	4	4	0	4
200-500	4	1	1	1	1

Figure 20: Number of data associated with the skill table.

Weak

Southern Gulf

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	1.17	1.80	0.93	1.86	0.81
0-20	1.10	1.69	0.93	2.05	0.76
20-50	1.61	1.94	0.93	[illegible]	3.00
50-100	3.18	3.23		2.67	

Cabot

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	0.53	0.85	[illegible]	[illegible]	0.52
0-20	[illegible]		[illegible]	[illegible]	[illegible]
20-50	0.64	0.80	[illegible]	0.70	0.57
50-100	[illegible]	0.82	0.50	0.83	[illegible]
100-200	1.17	1.32	0.76	1.14	1.20
200-500	0.57	0.55	0.61	0.75	[illegible]

Halifax

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	0.69	0.61	0.80	2.54	[illegible]
0-20	0.70	0.73	0.68	131.04	0.42
20-50	0.79	0.62	0.90	1.38	0.81
50-100	[illegible]	[illegible]	[illegible]	1.33	[illegible]
100-200	1.10	1.04	1.27		1.03
200-500	4.30	2.65	12.49	2.50	8.71

Strong

Southern Gulf

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	0.99	1.86	0.99	0.93	0.76
0-20	1.00	1.88	0.97	0.99	0.76
20-50	0.82		1.26	[illegible]	
50-100	0.51	0.51			

Cabot

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	[illegible]	[illegible]	[illegible]	[illegible]	0.71
0-20	[illegible]		[illegible]	[illegible]	3.35
20-50	[illegible]	[illegible]	[illegible]	0.60	0.57
50-100	[illegible]	[illegible]	[illegible]	0.54	[illegible]
100-200	1.04	1.00	0.76	0.75	1.50
200-500	0.66	0.95	1.13	[illegible]	[illegible]

Halifax

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	[illegible]	[illegible]	0.59	1.77	[illegible]
0-20	[illegible]	[illegible]	0.52		[illegible]
20-50	0.54	[illegible]	0.81	1.82	0.55
50-100	[illegible]	[illegible]	[illegible]	1.69	[illegible]
100-200	1.49	1.06			3.28
200-500	4.30	2.65	12.49	2.50	8.71

Weak

Southern Gulf

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	32	9	6	11	6
0-20	20	4	4	7	5
20-50	8	2	2	3	1
50-100	4	3	0	1	0

Cabot

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	79	17	16	22	24
0-20	10	0	3	4	3
20-50	24	7	3	5	9
50-100	16	3	4	6	3
100-200	15	4	2	3	6
200-500	14	3	4	4	3

Halifax

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	103	33	24	11	35
0-20	22	9	4	1	8
20-50	28	7	8	4	9
50-100	30	9	7	4	10
100-200	15	6	3	0	6
200-500	8	2	2	2	2

Strong

Southern Gulf

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	16	4	4	3	5
0-20	13	3	3	2	5
20-50	2	0	1	1	0
50-100	1	1	0	0	0

Cabot

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	46	7	12	13	14
0-20	7	0	3	3	1
20-50	11	2	2	2	5
50-100	13	2	4	5	2
100-200	9	2	2	2	3
200-500	6	1	1	1	3

Halifax

Depth (m)	Annual	Winter	Spring	Summer	Fall
ALL	65	22	14	6	23
0-20	13	5	3	0	5
20-50	19	5	5	2	7
50-100	16	5	4	2	5
100-200	9	5	0	0	4
200-500	8	2	2	2	2

Figure 21: Comparison of model skill for weak versus strong criteria for the Southern GSL, CS and Hfx boxes, including the data counts.

5.4 Model Validation: Ice Model

The CANOPA ice (LIM2) model was validated with Gulf of Saint Lawrence data compiled at BIO for the last 5 decades. The data were derived from 2 sources. One source is the 1963-1998 coarse resolution (0.5° latitude X 1.0° longitude) data made from Canadian Ice Service (CIS) ice charts commonly refered to as the Drinkwater et at. (1999) database. The second source was a 1998-2011 BIO compilation of CIS digitized high resolution (4km) charts (Pers. Comm. R. Pettipas, 2011).

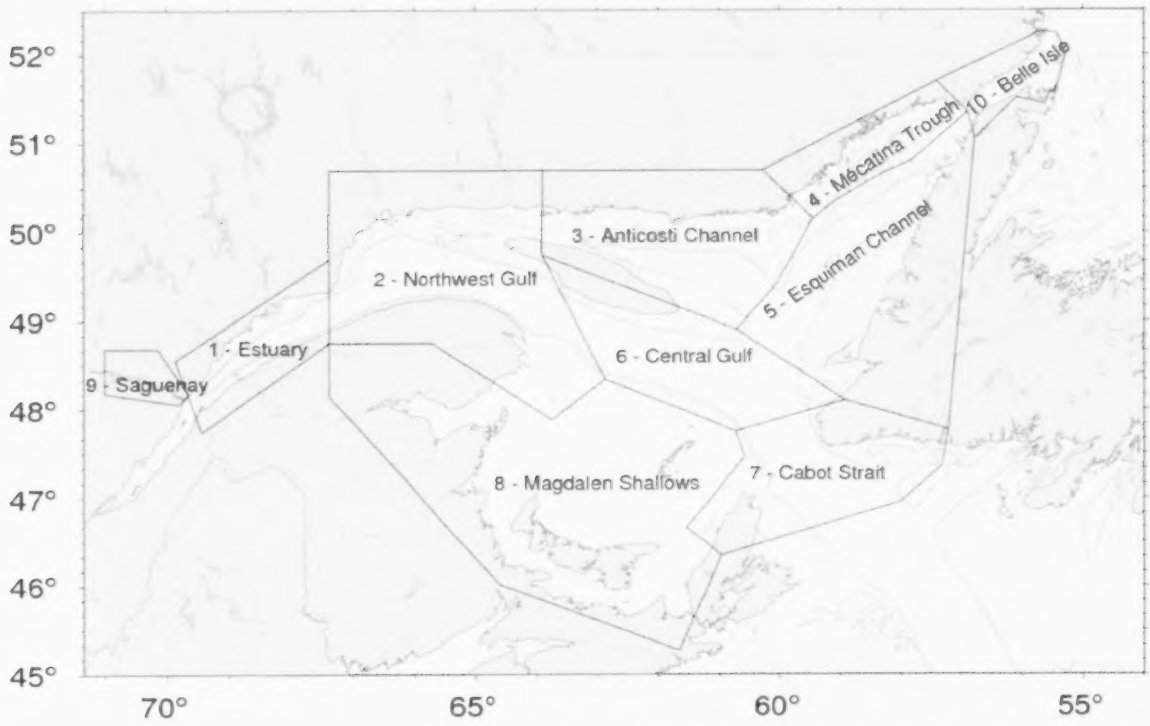


Figure 22: Definition of Gilbert Boxes (reproduced from Galbraith et al. 2011)

The approach taken for the validation involves computing statistics for distinct regions of the Gulf refered to as Gilbert Boxes (Figure 22). Box 9, Saguenay region is excluded from the analysis due to insuficient resolution in model and data. The statistics chosen for this part of the validation were day of first appearance, day of last appearance and peak volume. The peak volume captures the net effect of concentration and thickness and is a good metric for capturing the intensity of the ice in a particular year. The day of first and last appearance define the timing of the ice season. The data provides 49 instances of the 3 statistics while the model provides only 1 from the Normal Year simulation. To provide a sense of the inherent variability, we compare the 49 year timeseries to the model climatological values. The result for the 9 regions are shown in Figures 23-31.

Results vary from region to region. The result for the entire Gulf is shown in Figure 32. It seems the model peak volume is a good fit through the highly variable time series. However, the ice seems to arrive about 10 days too early and leave 20 days too late. Figure 33 shows the area weighted average for the the 9 regions. As can be seen, the model compares favourably to the area weighted data. In particular, averaging the data has significantly improved the timing error. Summary plots of the 3 ice statistics are shown in figures 34, 35, 36 and 37. The summaries show the means $\pm$ 1 standard deviation for all the regions. First appearance tends to be about 2 weeks too early for the Estuary and Northwest Gulf (outside the error bars by about 2 stds), but is within the error bars for the remaining regions and the region average. For the last appearance, the Estuary and Central Gulf is just slightly outside the error bars (too late) but Mecatina Trough, Esquiman Channel and Belle Isle are between 2 and 3 std's too early. Interestingly, the model does a good job simulating the average for all regions, indicating regions where we are too early offset the late regions. The peak volume, best seen in the normalized version of the graph, shows that we are doing reasonably well (within error) for most regions and the average. The model is somewhat underestimating Mecatina Trough and Esquiman Channel (just slightly outside error bars) but severely underestimating Belle Isle (4 std error) or only 10 percent of expected volume modelled. It is believed that the regions of underestimation of volume are adversly affected by the boundary conditions. The Esquiman region (and to a large extent Central and Cabot regions) is affected by the warm water inflow from the eastern boundary that flows through the eastern part of CS along the Newfoundland coastline (recall the discussion in section 4), while the Belle Isle (and possibly Mecatina) region, are poorly modelled because the model lacks ice inflow boundary conditions at Belle Isle. It is observed that the model is ice free in the cells immediately adjacent to the SBI open boundary. This is explained as being due to inward advection of > 0 deg water through the open boundary which limits ice formation to farther downstream of the boundary. Changes to the winter open boundary conditions are being considered.

A relevant question is the degree to which the lack of ice flux at the SBI open boundary affects the model estimate of the total GSL ice volume. An estimate of the ice flux through SBI from the CECOM circulation model (Dr. Y. Wu, BIO, Pers. Comm.) indicates about 1.5–2.0 km^3 total from January to March, much smaller than the average peak ice volume of ~ 60 km^3 (figure 6).

Another estimate can be derived from the ice volume data for the Belle Isle region. Consider the SBI box to be like a conveyor belt of length L with ice flowing in upstream and exiting downstream with characteristic velocity V. Reasonable values for L and V are 100-200km and 5 cm/s which means that the "evacuation" ratio $(V\Delta t)/L \sim 1$ (as opposed to $\gg 1$). This means that (approximately) in a month all the ice originally in the box would be advected out and any ice entering the box would not exit the box. These assumptions imply that the standing stock of ice (e.g. total monthly ice volume) serves as an upper bound for the flux of ice into the SBI box. This is strictly true for

the ice growth phase, and also applicable if we assume that the ice decay rate is spatially uniform. From figure 39, we estimate this to be about 5 km^3, which is also much smaller than the average peak ice volume of ~ 60 km^3.

Our conclusion is that ignoring the ice flux into the SBI open boundary should not greatly affect the model estimate of the total ice volume in the GSL. Including it should improve our model-data comparison in boxes near the northeast open boundary.

A final comparison involves looking at monthly averages of the peak ice volumes for the regions discussed above. Figure 38 and 39 show the volumes for the 9 Gilbert boxes as well as the for the sum of all the regions. As before, the standard deviation departure is included on the graph and taken to represent the error. Overall for the sum of all regions the model matches the data almost perfectly. The 4 regions in Figure 38 which are the more westerly/northwesterly parts of the Gulf, represent the regions with better model agreement, although the model has a tendency to overestimate the data. Regions shown in Figure 39 which are more easterly and central areas tend not to do as well. These are the same regions that had problems with volume due to the nearness of the model open boundary condition. The tendency is to understimate the peak ice volumes in these region by about 1 std but occasionally more as in the Belle Isle box.

6 Summary

This report has described the development and validation of a NEMO-OPA numerical circulation model ("CANOPA" model) for use on a domain that includes the shelf seas of Maritime Canada.

Section 2 describes the model code and changes made by the authors related to its use as a shelf model, while section 3 provides details of the model set up.

Section 4 discusses the philosophy behind model validation and presents some basic model validation results. Model transport timeseries are compared to historical data at 2 locations on the Scotian Shelf and are found to follow the data within error bars (figure 4). The model ice field is found to qualitatively follow the characteristic growth and decay pattern in the GSL (figure 5). The monthly timeseries of total ice volume is within error bars of the data (figure 6), with the fit being better during the ice growth phase and underestimating the data during the ice decay phase.

Section 5 contains quantitative comparisons between model currents and current meter data, and the model's ice simulation versus 49 years of ice data. For the velocity field, a vector-by-vector comparison showed that the distribution of model-minus-observation current speeds (figure 14) was Gaussian-like with a median of 0.003 m/s and a mean of 0.0205 m/s, i.e. slightly skewed toward overestimation by the model. In terms of speed ratio (figure 15), a log-normal distribution was approximated with median and mean of 1.06, with $\sim 90\%$ of the ratios within one order of magnitude. Vector difference

calculations (figures 12 and 16) showed the most common error magnitude was 2cm/s, with a mean error angle of -4 degrees. The distribution of speeds for the entire domain was well approximated by the model (figure 17).

The current meter data were grouped in time and space to form a seasonal climatology, and the domain broken into 8 regions for comparison purposes (figure 18). The ratio r of error kinetic energy to observation kinetic energy, was computed on a regional basis for various depth intervals. The skill over all regions, seasons and layers is 1.05, considered reasonable for such a large region. On a layer basis, (for all regions and seasons), there is fair skill ($0.5 < r < 1.0$) for 0-50m and 500-2000m with best agreement below 2000m. For seasonal differences for all the regions combined, we find that fall is the season with best skill with only the 100-500 lacking good ($r < 0.5$) or fair skill.

The ice comparison was based on the Gilbert boxes (figure 22). On an area-weighted basis (figures 33–36) the peak ice volume and day of last appearance were well simulated by the model, while the average day of first arrival was ~ 10 days too early. In general we found that, for all metrics, the model performed better for western boxes which are isolated from open boundary effects, although exceptions exist (figures 34–36). Regarding the unmodelled flux of ice into the Strait of Belle Isle, we showed this is small compared to the total ice volume, and that the omission of this open boundary condition can explain the degraded model performance in the SBI and Mecatina boxes.

7 Acknowledgements

The authors would like to thank Dr. Z. Wang (BIO) and B. deTracey (BIO) for their reviews of this document. The DFO Aquatic Invasive Species program and the DFO Centre for Ocean Model Development for Applications (COMDA) provided support for model development.

8 References

Anderson, C. and P.C. Smith. 1989. Oceanographic observations on the Scotian Shelf during CASP. Atmosphere-Ocean. **29(1)**, 130-156.

Brickman, D. and A. Drozdowski, 2012. Atlas of Model Currents and Variability in Maritime Canadian Waters. Can. Tech. Rep. Hydrogr. Ocean Sci. 277: vii + 64 pp.

Chaffey, J.D. and D.A. Greenberg, 2003. resolute: A Semi-Automated Finite Element Mesh Generation Routine. Can. Tech. Rep. Hydrogr. Ocean Sci. 225: vi + 33 pp.

Drinkwater, K. F., R.G. Pettipas, G.L. Bugden, and P. Langille. 1999. Climatic data for the Northwest Atlantic: a sea ice database for the Gulf of St. Lawrence and the Scotian Shelf. Can. Tech. Rep. Hydrogr. Ocean Sci. 199: iv + 134 p.

Dupont, F., C. G. Hannah, D. A. Greenberg, J. Y. Cherniawsky and C. E. Naimie. 2002. Modelling System for Tides. Can. Tech. Rep. Hydrogr. Ocean Sci. 221: vii + 72 pp. ii

ETOPO2v2 Global Gridded 2-minute Database, National Geophysical Data Center, National Oceanic and Atmospheric Administration, U.S. Dept. of Commerce, http://www.ngdc.noaa.gov/mgg/global/etopo2.html.

Flather, R. A. 1976. A tidal model of the northwest European continental shelf. Memo. Soc. Roy. Sci. Liege, 6 (10), 141-164.

Galbraith, P.S., Chass, J., Gilbertt, D., Larouche, P. Brickman, D., Pettigrew, B., Devine, L., Gosselin, A., Pettipas, R.G. and Lafleur, C. 2011. Physical Oceanographic Conditions in the Gulf of St. Lawrence in 2010. DFO Can. Sci. Advis. Sec. Res. Doc. 2011/045. iv + 82 p.

Gregory, D.N. 2004. Ocean Data Inventory (ODI): A database of ocean current, temperature and salinity time series for the Northwest Atlantic. CSAS Res. Doc. 2004/097.

Hannah, C.G., Shore, J.A., Loder, J.W. and Naimie, C.E. 2001. Seasonal circulation on the western and central Scotian Shelf. J. Phys. Ocean. V.31, 591-615.

Large, W. and S. Yeager. 2004. Diurnal to decadal global forcing for ocean and sea-ice models: the data sets and flux climatologies. CGD Division of the National Center for Atmospheric Research, NCAR Technical Note: NCAR/TN-460+STR.

Large, W., J. McWilliams, and S. Doney. 1994. Oceanic vertical mixing: A review and a model with nonlocal boundary layer parameterization, Rev. Geophys., 32, 363-403.

Le Fouest, V., B. Zakardjian, F. J. Saucier, and M. Starr. 2005. Seasonal versus synoptic variability in planktonic production in a high-latitude marginal sea: The Gulf of St. Lawrence (Canada), J. Geophys. Res., 110, C09012, doi:10.1029/2004JC002423.

Loder, J.W., Hannah, C.G., Petrie, B.D., and Gonzalez, E.A. 2003. Hydrographic and transport variability on the Halifax section. J. Geophys. Res. **108** (C11), 8003, doi:10.1029/2001JC001267.

Lyard, F., F. Lefevre, T. Letellier and O. Francis. 2006. Modelling the global ocean tides: modern insights from FES2004. Ocean Dynamics **56**, 394-415.

Madec G. 2008. "NEMO ocean engine". Note du Pole de modlisation, Institut Pierre-Simon Laplace (IPSL), France, No 27 ISSN No 1288-1619.

Madec, G., Delecluse P., Imbard M., and Lvy C. 1998. LIM2 ice model, in OPA Reference manual: OPA 8.1 Ocean General Circulation Model reference manual. Note du Pole de modlisation, Institut Pierre-Simon Laplace (IPSL), France, No11, 91pp.

Mellor, G. L. and T. Yamada, Development of a turbulence closure model for geophysical fluid problems. 1982. Rev. Geophys. and Space Phys., 20(4), 851-875.

NetCDF, Network common data form.
http://www.unidata.ucar.edu/software/netcdf/docs/

Rodean, H. C. 1996. Stochastic Lagrangian Models of Turbulent Diffusion, Meteor. Monogr., No. 48, Amer. Meteor. Soc., 84 pp.

Sheng, J. Wright, D.G., Greatbatch, R.J., Dietrich, D.E. 1998. CANDIE: A new version of the DieCast ocean circulation model. Journal of Atmospheric and Oceanic Technology **15**, 1414-1432.

Smagorinsky, J. 1963. General Circulation experiments with the primitive equations I. The basic experiment. Monthly Weather Review, **91(3)**, 99-164.

Smith, P.C. 1983. The mean and seasonal circulation off southwest Nova Scotia. J. Phys. Oceanog. **13(6)**, 1034-1054.

Figure 23: Time series of 49 years of ice conditions in Gilbert Box 1. Horizontal line denotes the CANOPA normal year prediction.

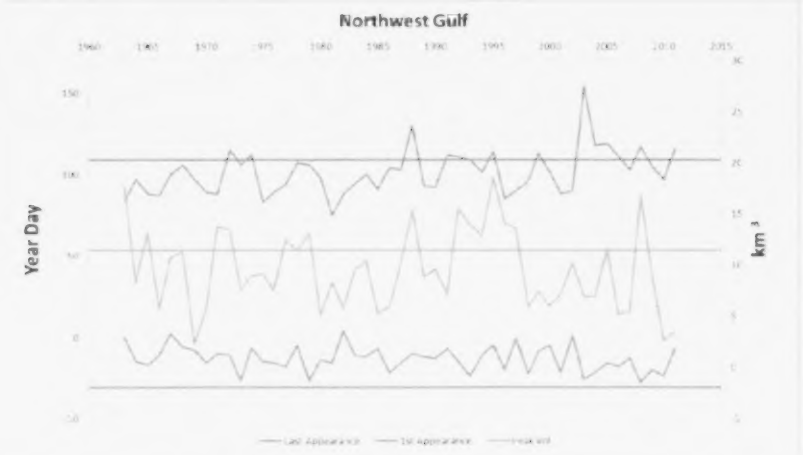


Figure 24: Time series of 49 years of ice conditions in Gilbert Box 2. Horizontal line denotes the CANOPA normal year prediction.

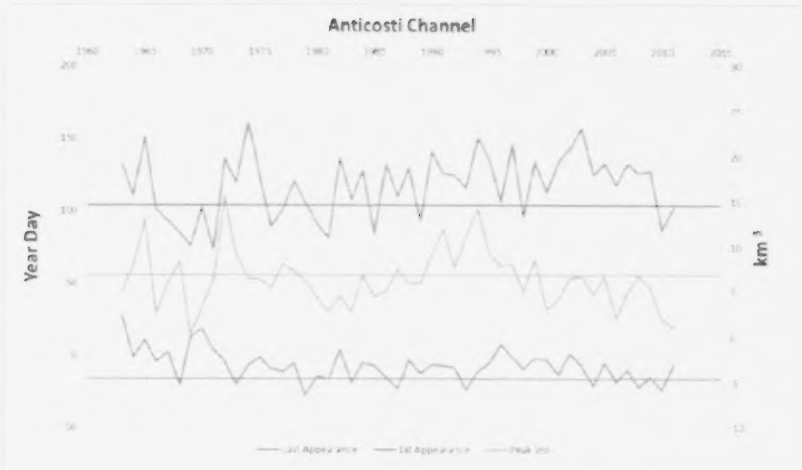


Figure 25: Time series of 49 years of ice conditions in Gilbert Box 3. Horizontal line denotes the CANOPA normal year prediction.

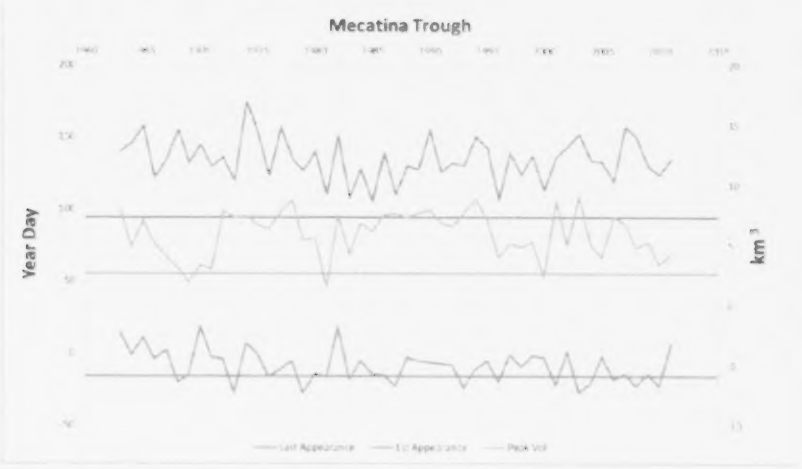


Figure 26: Time series of 49 years of ice conditions in Gilbert Box 4. Horizontal line denotes the CANOPA normal year prediction.

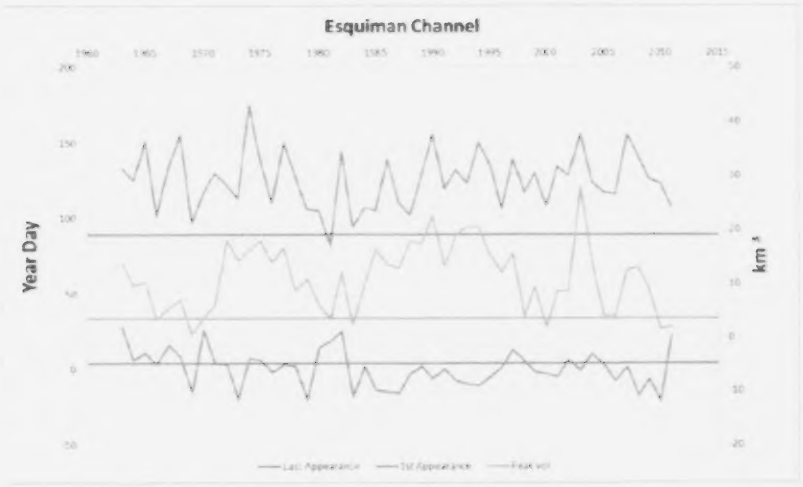


Figure 27: Time series of 49 years of ice conditions in Gilbert Box 5. Horizontal line denotes the CANOPA normal year prediction.

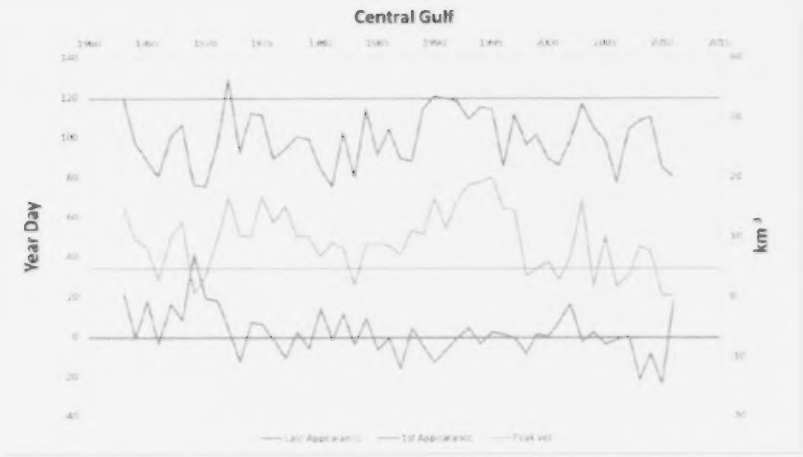


Figure 28: Time series of 49 years of ice conditions in Gilbert Box 6. Horizontal line denotes the CANOPA normal year prediction.

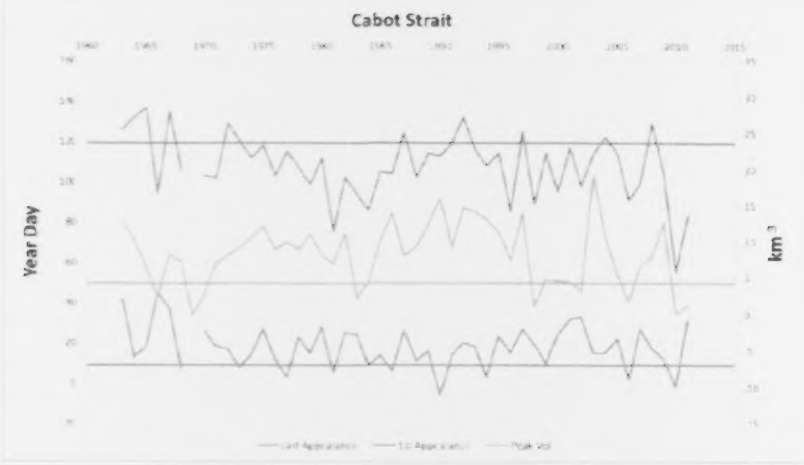


Figure 29: Time series of 49 years of ice conditions in Gilbert Box 7. Horizontal line denotes the CANOPA normal year prediction.

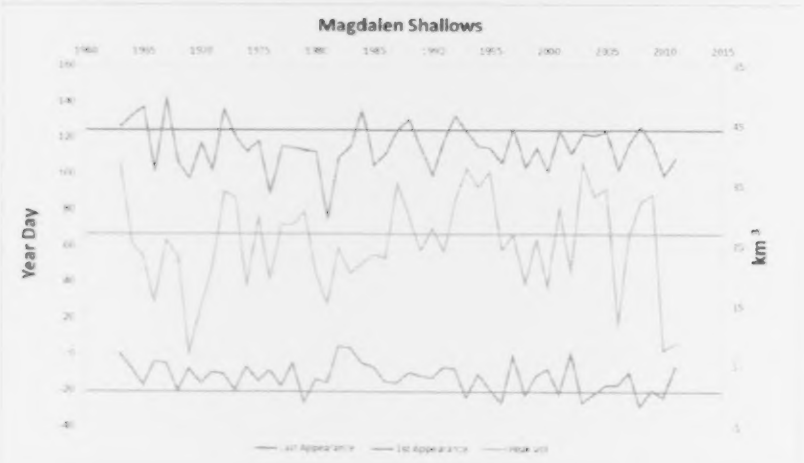


Figure 30: Time series of 49 years of ice conditions in Gilbert Box 8. Horizontal line denotes the CANOPA normal year prediction.

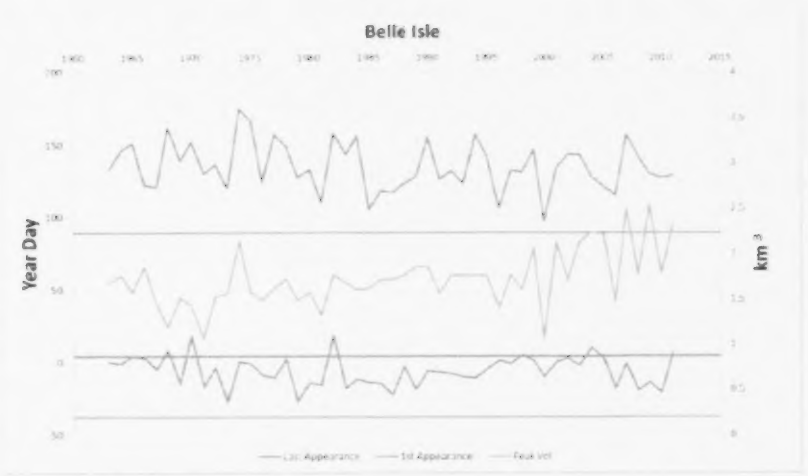


Figure 31: Time series of 49 years of ice conditions in Gilbert Box 10. Horizontal line denotes the CANOPA normal year prediction.

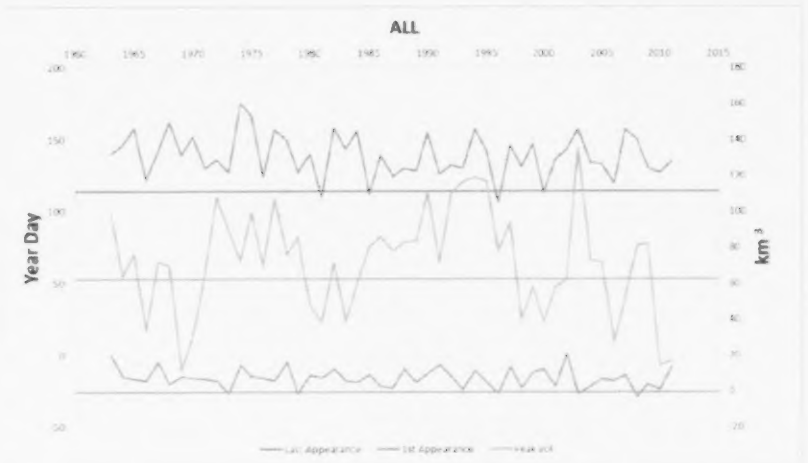


Figure 32: Time series of 49 years of ice conditions in the entire Gulf of Saint Lawrence. Horizontal line denotes the CANOPA normal year prediction.

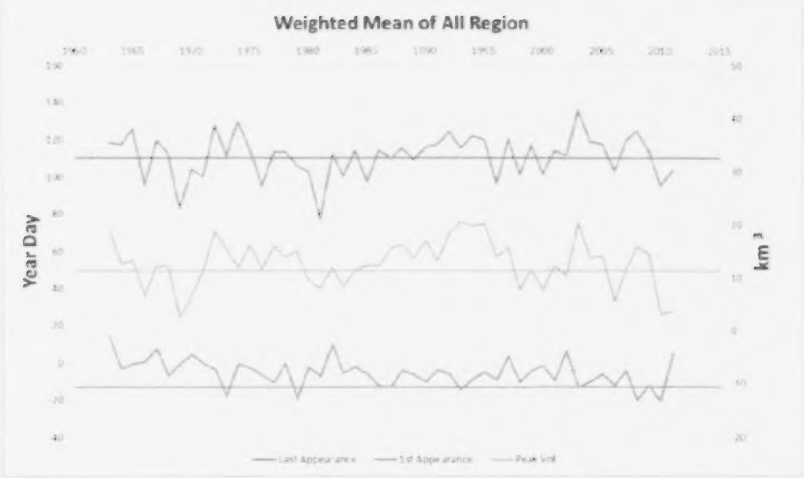


Figure 33: Time series of 49 years of regionally averaged (weighted by area) ice conditions in the Gulf of Saint Lawrence.

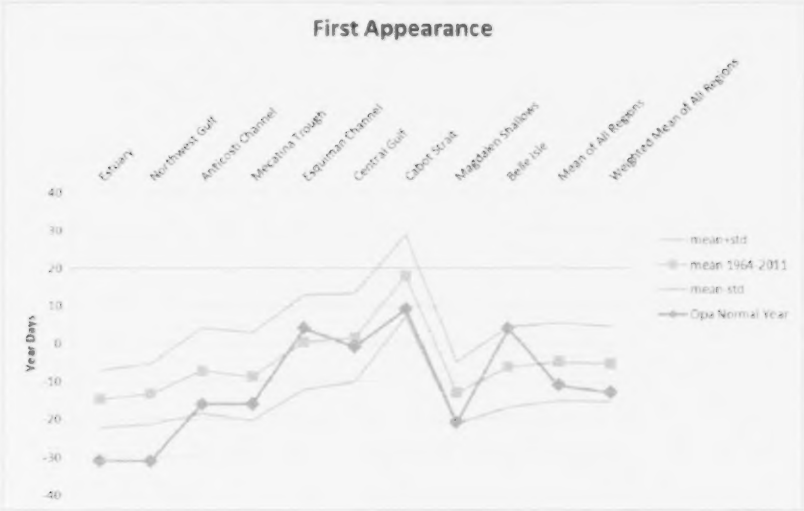


Figure 34: Regional 49 year mean of first appearance and CANOPA normal year value. Included are lines showing standard deviation departure from the mean of the data.

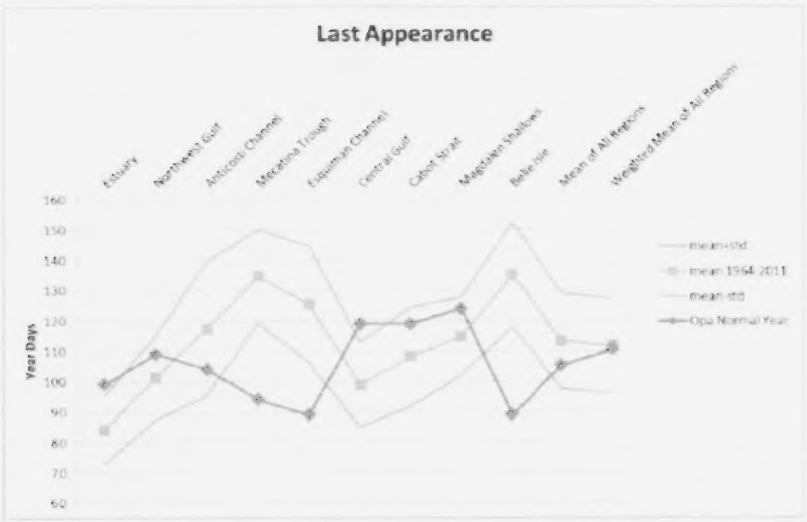


Figure 35: Regional 49 year mean of last appearance and CANOPA normal year value. Included are lines showing standard deviation departure from the mean of the data.

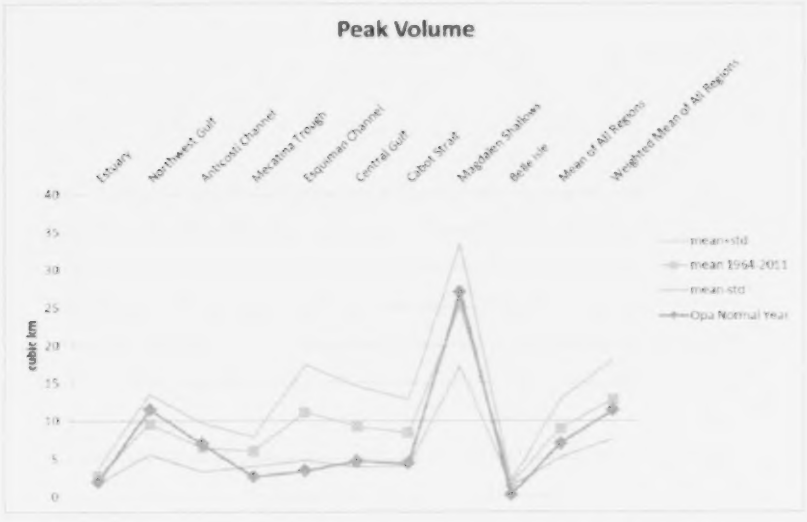


Figure 36: Regional 49 year mean of peak volume and CANOPA normal year value. Included are lines showing standard deviation departure from the mean of the data.

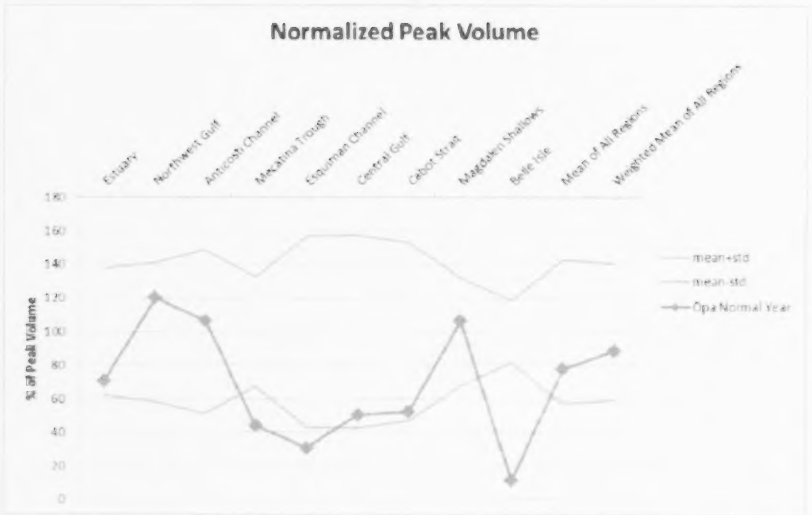


Figure 37: Regional 49 year mean of normalized peak volume and CANOPA normal year value. Included are lines showing standard deviation departure from the mean of the data. All values normalized by peak volume of data.

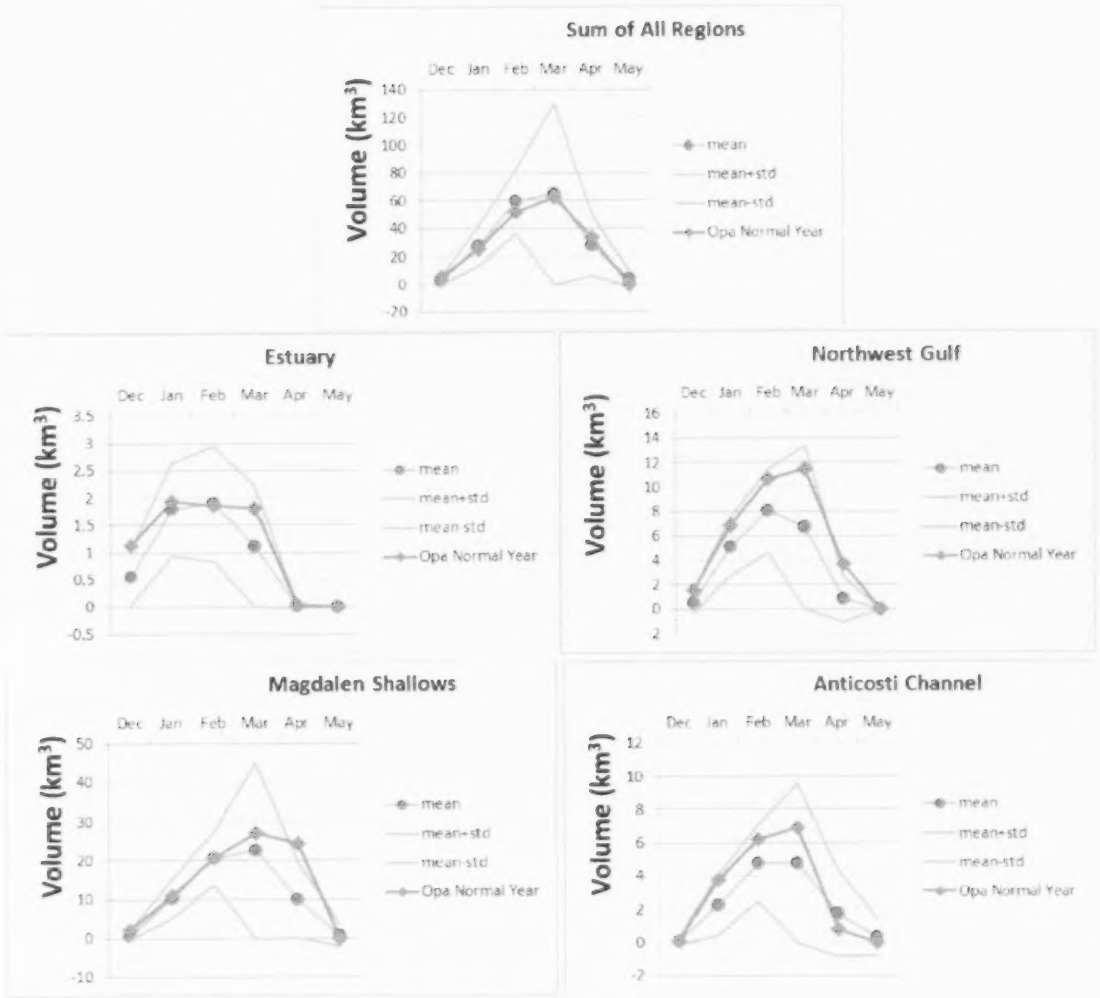


Figure 38: Monthly mean peak ice volume based on 49 years of data and CANOPA normal year. Included are lines showing standard deviation departure from the mean of the data. Part 1: Sum of all regions and western/north-western regions.

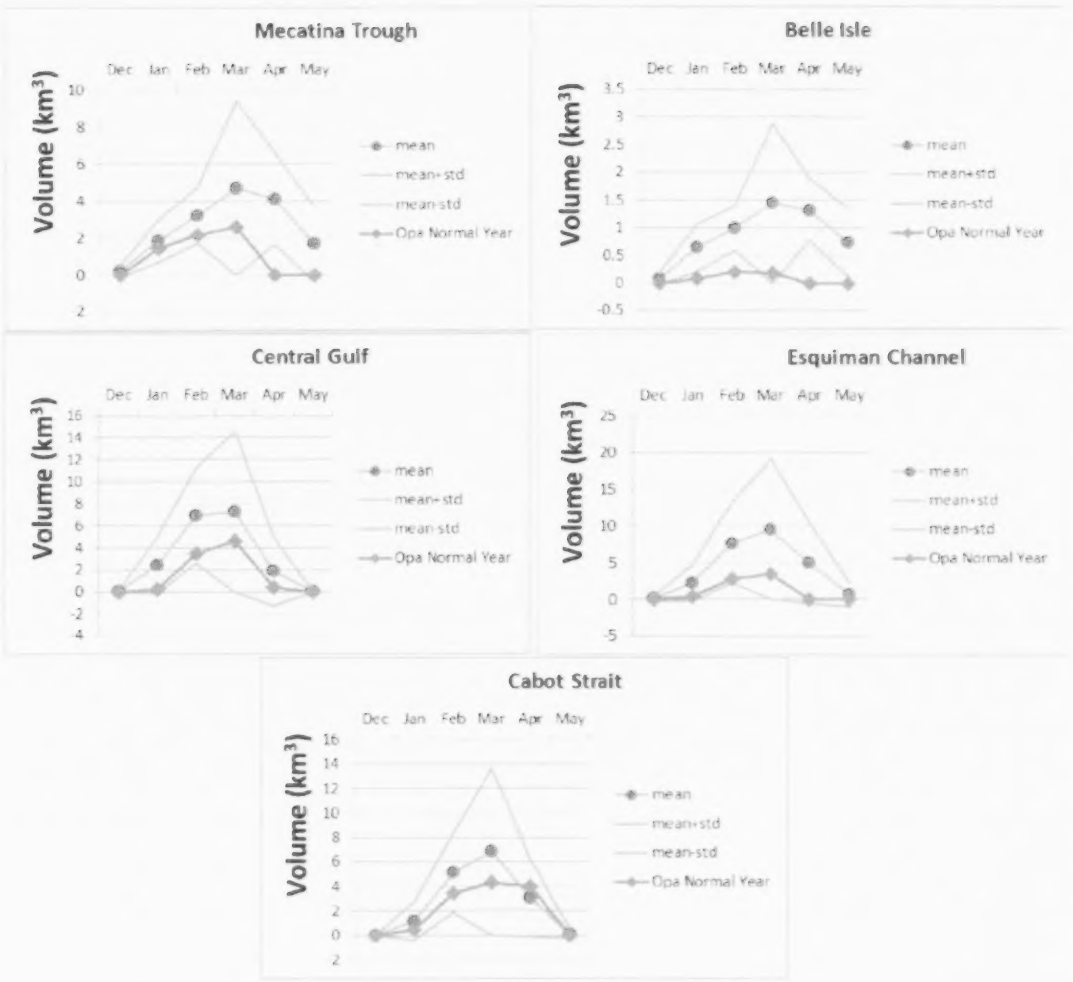


Figure 39: Monthly mean peak ice volume based on 49 years of data and CANOPA normal year. Included are lines showing standard deviation departure from the mean of the data. Part 2 Eastern and Central Regions

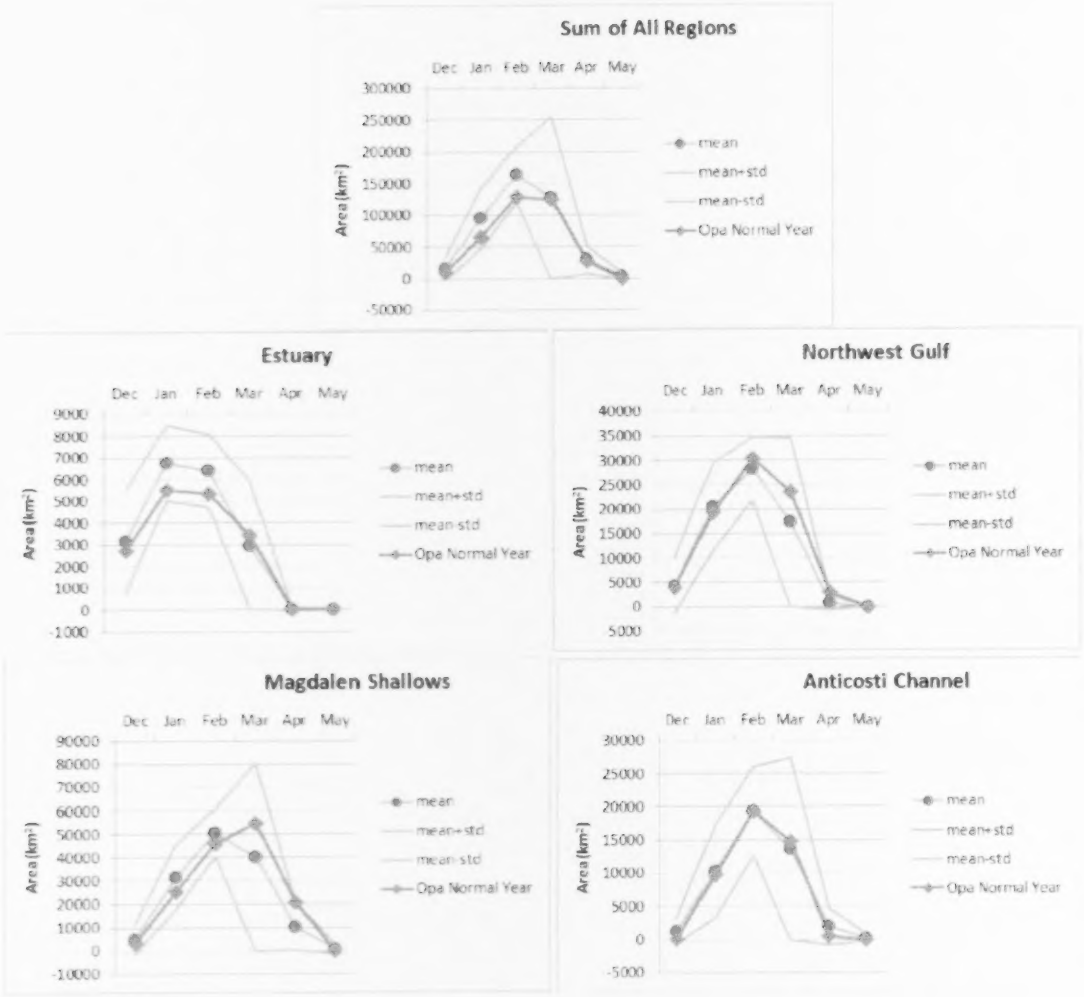


Figure 40: Monthly mean ice area based on 49 years of data and CANOPA normal year. Included are lines showing standard deviation departure from the mean of the data. Part 1: Sum of all regions and western/north-western regions.

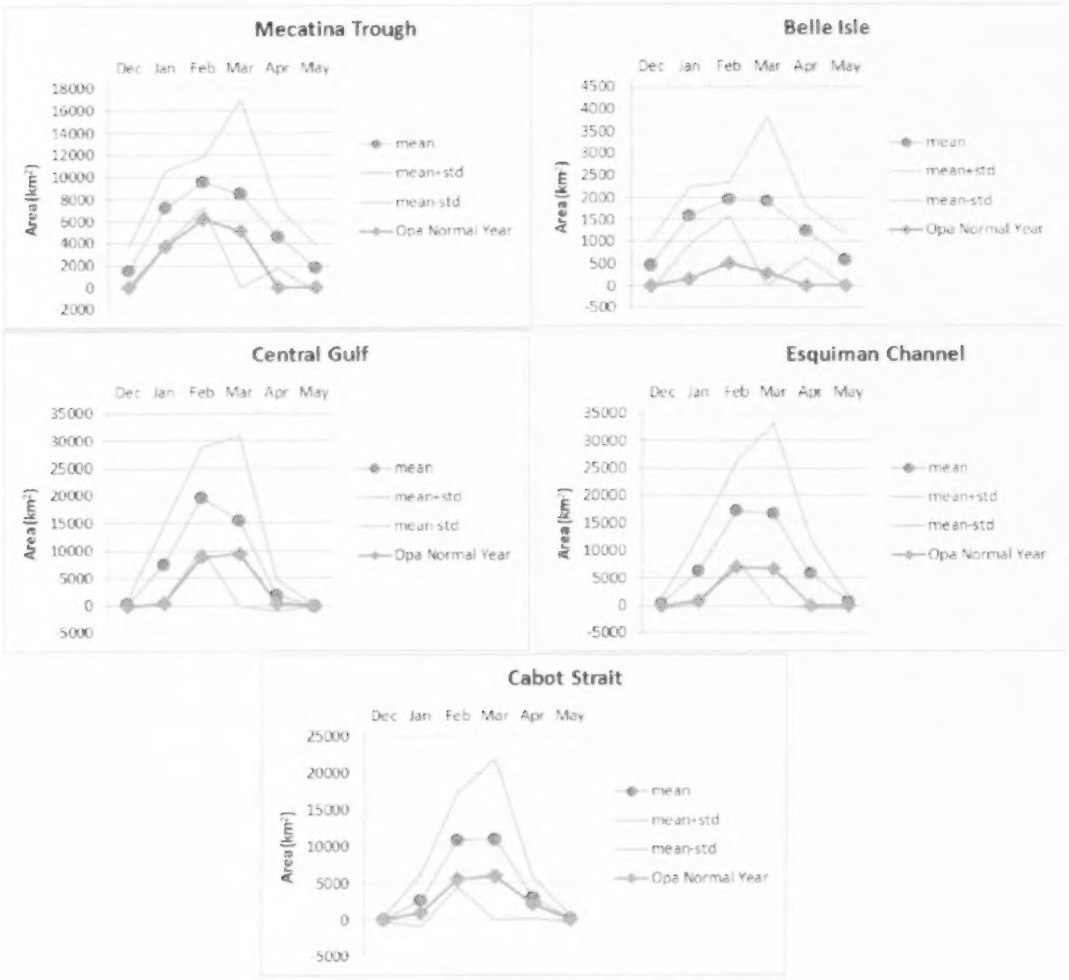


Figure 41: Monthly mean ice area based on 49 years of data and CANOPA normal year. Included are lines showing standard deviation departure from the mean of the data. Part 2 Eastern and Central Regions